

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 02:19:50 ; Search time 2768 Seconds  
(without alignments)  
3690.421 Million cell updates/sec

Title: US09897438BK-2  
Perfect score: 351  
Sequence: 1 GARCARTGYGNACNATNAT.....AYCARGCNTGTGGGNCYTN 351

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	16	4.6	23	6	AR136101	Sequence AR136101 Sequence
2	16	4.6	885	6	AX047947	Sequence AX047947 Sequence
3	16	4.6	1044	6	AX047954	Sequence AX047954 Sequence
4	16	4.6	2379	6	AX211531	Sequence AX211531 Sequence
5	15	4.3	426	6	AX055062	Sequence AX055062 Sequence
6	15	4.3	696	6	AX255037	Sequence AX255037 Sequence
7	15	4.3	729	6	AX255040	Sequence AX255040 Sequence
8	15	4.3	987	6	AX068381	Sequence AX068381 Sequence
9	15	4.3	1155	6	AX108644	Sequence AX108644 Sequence
10	15	4.3	2019	6	AX179624	Sequence AX179624 Sequence
11	15	4.3	2277	6	AR004980	Sequence AR004980 Sequence
12	14	4.0	20	6	I26713	Sequence I26713 Sequence 14
13	14	4.0	1026	6	AX350993	Sequence AX350993 Sequence
14	14	4.0	1134	6	AX172419	Sequence AX172419 Sequence
15	14	4.0	1584	6	AX167244	Sequence AX167244 Sequence
16	14	4.0	1629	6	AX167247	Sequence AX167247 Sequence
17	14	4.0	1695	6	AX350990	Sequence AX350990 Sequence
18	14	4.0	1728	6	AX237003	Sequence AX237003 Sequence
19	14	4.0	2094	6	AX350978	Sequence AX350978 Sequence
20	14	4.0	2094	6	AX360312	Sequence AX360312 Sequence
21	13	3.7	20	6	A23188	Sequence A23188 Artificial
22	13	3.7	20	6	AR050672	Sequence AR050672 Sequence
23	13	3.7	20	6	E08318	Sequence E08318 Probe for d
24	13	3.7	23	6	AR009221	Sequence AR009221 Sequence
25	13	3.7	23	6	I49042	Sequence I49042 Sequence 13
26	13	3.7	23	6	I73415	Sequence I73415 Sequence 13
27	13	3.7	51	6	AX356144	Sequence AX356144 Sequence
28	13	3.7	72	6	AR042415	Sequence AR042415 Sequence
29	13	3.7	351	6	AX154583	Sequence AX154583 Sequence
30	13	3.7	390	6	AX193710	Sequence AX193710 Sequence
31	13	3.7	693	6	AX151702	Sequence AX151702 Sequence
32	13	3.7	693	6	AX193707	Sequence AX193707 Sequence
33	13	3.7	696	6	AX280162	Sequence AX280162 Sequence
34	13	3.7	729	6	AX350987	Sequence AX350987 Sequence
35	13	3.7	747	6	AX193684	Sequence AX193684 Sequence
36	13	3.7	804	6	AX398327	Sequence AX398327 Sequence
37	13	3.7	882	6	AX398330	Sequence AX398330 Sequence
38	13	3.7	906	6	AX398703	Sequence AX398703 Sequence
39	13	3.7	1098	6	AX163791	Sequence AX163791 Sequence
40	13	3.7	1602	6	AR150957	Sequence AR150957 Sequence
41	13	3.7	1644	6	AX269154	Sequence AX269154 Sequence
42	13	3.7	2082	6	AR166376	Sequence AR166376 Sequence
43	13	3.7	2109	6	AX350975	Sequence AX350975 Sequence
44	13	3.7	2214	6	AX350981	Sequence AX350981 Sequence
45	13	3.7	2217	6	AX364578	Sequence AX364578 Sequence

ALIGNMENTS

RESULT 1	AR136101	Sequence 35 from patent US 6136581.	23 bp	DNA	linear	PAT 16-JUN-2001
LOCUS	AR136101	Sequence 35 from patent US 6136581.				
DEFINITION	AR136101	Sequence 35 from patent US 6136581.				
ACCESSION	AR136101	Sequence 35 from patent US 6136581.				
VERSION	AR136101.1	GI:14476773				
KEYWORDS						
SOURCE		Unknown.				
ORGANISM		Unknown.				
REFERENCE		Unclassified.				
AUTHORS		1 (bases 1 to 23)				
TITLE		Joho,K.E. and Plowman,G.D.				
JOURNAL		Kinase genes and uses				
FEATURES		Patent: US 6136581-A 35 24-OCT-2000;				
		Location/Qualifiers				

```
source 1. .23
/organism="unknown"
BASE COUNT 3 a 1 c 8 g 4 t 7 others
ORIGIN

Query Match 4.6%; Score 16; DB 6; Length 23;
Best Local Similarity 100.0%; Pred. No. 7.3e-21;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 322 GCGARGTNTAYGARG 337
|||||
Db 7 GCGARGTNTAYGARG 22

RESULT 2
AX047947
LOCUS AX047947 885 bp DNA linear PAT 15-DEC-2000
DEFINITION Sequence 14 from Patent WO0070045.
ACCESSION AX047947
VERSION AX047947.1 GI:11876870
KEYWORDS
SOURCE .
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 885)
AUTHORS Barclay,A.N., Brown,M.H., Gorman,D.M., Lanier,L.L., Wright,G.J.,
Cherwinski,H., Phillips,J.H., Hoek,R.M. and Sedgwick,J.D.
TITLE Ox2 receptor homologs
JOURNAL Patent: WO 0070045-A 14 23-NOV-2000;
MEDICAL RESEARCH COUNCIL (GB) ; SCHERING CORPORATION (US)
FEATURES
source Location/Qualifiers
1. .885
/db_xref="taxon:32644"
BASE COUNT 182 a 98 c 133 g 114 t 358 others
ORIGIN

Query Match 4.6%; Score 16; DB 6; Length 885;
Best Local Similarity 100.0%; Pred. No. 1.5e-21;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 229 AAYGTNWSNACNGTNA 244
|||||
Db 697 AAYGTNWSNACNGTNA 712

RESULT 3
AX047954
LOCUS AX047954 1044 bp DNA linear PAT 15-DEC-2000
DEFINITION Sequence 21 from Patent WO0070045.
ACCESSION AX047954
VERSION AX047954.1 GI:11876877
KEYWORDS
SOURCE .
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1044)
AUTHORS Barclay,A.N., Brown,M.H., Gorman,D.M., Lanier,L.L., Wright,G.J.,
Cherwinski,H., Phillips,J.H., Hoek,R.M. and Sedgwick,J.D.
TITLE Ox2 receptor homologs
JOURNAL Patent: WO 0070045-A 21 23-NOV-2000;
MEDICAL RESEARCH COUNCIL (GB) ; SCHERING CORPORATION (US)
FEATURES
source Location/Qualifiers
1. .1044
/db_xref="taxon:32644"
BASE COUNT 220 a 117 c 154 g 131 t 422 others
ORIGIN

Query Match 4.6%; Score 16; DB 6; Length 1044;
Best Local Similarity 100.0%; Pred. No. 1.4e-21;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 229 AAYGTNWSNACNGTNA 244
|||||
Db 697 AAYGTNWSNACNGTNA 712

RESULT 4
AX211531
LOCUS AX211531 2379 bp DNA linear PAT 07-SEP-2001
DEFINITION Sequence 12 from Patent WO0159112.
ACCESSION AX211531
VERSION AX211531.1 GI:15523810
KEYWORDS
SOURCE .
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 2379)
AUTHORS Holloway,J.L. and Sheppard,P.O.
TITLE Anti-angiotensin intestinal peptides, zdist5
JOURNAL Patent: WO 0159112-A 12 16-AUG-2001;
ZymoGenetics, Inc. (US)
FEATURES
source Location/Qualifiers
1. .2379
/organism="synthetic construct"
/db_xref="taxon:32630"
BASE COUNT 270 a 286 c 502 g 272 t 1049 others
ORIGIN

Query Match 4.6%; Score 16; DB 6; Length 2379;
Best Local Similarity 100.0%; Pred. No. 9.7e-22;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 123 NWSNGGNWSNTGYMGN 138
|||||
Db 1512 NWSNGGNWSNTGYMGN 1527

RESULT 5
AX055062
LOCUS AX055062 426 bp DNA linear PAT 13-JAN-2001
DEFINITION Sequence 3 from Patent WO0073458.
ACCESSION AX055062
VERSION AX055062.1 GI:12228387
KEYWORDS
SOURCE .
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 426)
AUTHORS Conklin,D.C.
TITLE Secreted alpha-helical protein-31
JOURNAL Patent: WO 0073458-A 3 07-DEC-2000;
ZymoGenetics, Inc. (US)
FEATURES
source Location/Qualifiers
1. .426
/organism="synthetic construct"
/db_xref="taxon:32630"
misc_feature 1..426
/notes="Degenerate polynucleotide sequence of zalpha31"
/notes="n = A,T,C or G"
BASE COUNT 101 a 41 c 70 g 43 t 171 others
ORIGIN

Query Match 4.3%; Score 15; DB 6; Length 426;
Best Local Similarity 100.0%; Pred. No. 7e-19;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 263 CNGARGARGCNAARG 277
|||||
Db 77 CNGARGARGCNAARG 91

RESULT 6
AX255037
```



LOCUS AX255037 696 bp DNA linear PAT 10-OCT-2001  
DEFINITION Sequence 3 from Patent WO0170986.  
ACCESSION AX255037  
VERSION AX255037.1 GI:16074534  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct  
artificial sequences.  
REFERENCE 1 (bases 1 to 696)  
AUTHORS Konlin,D.C. and Presnell,S.R.  
TITLE Helical protein zalpha51  
JOURNAL Patent: WO 0170986-A 3 27-SEP-2001;  
ZymoGenetics, Inc. (US)  
FEATURES  
source  
1. .696  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="degenerate sequence"  
BASE COUNT 74 a 82 c 121 g 326 others  
ORIGIN  
Query Match 4.3%; Score 15; DB 6; Length 696;  
Best Local Similarity 100.0%; Pred. No. 5.7e-19;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 257 AYTNCNGARGARG 271  
|||||  
Db 449 AYTNCNGARGARG 463  
RESULT 7  
AX255040  
LOCUS AX255040 729 bp DNA linear PAT 10-OCT-2001  
DEFINITION Sequence 6 from Patent WO0170986.  
ACCESSION AX255040  
VERSION AX255040.1 GI:16074537  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct  
artificial sequences.  
REFERENCE 1 (bases 1 to 729)  
AUTHORS Konlin,D.C. and Presnell,S.R.  
TITLE Helical protein zalpha51  
JOURNAL Patent: WO 0170986-A 6 27-SEP-2001;  
ZymoGenetics, Inc. (US)  
FEATURES  
source  
1. .729  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="degenerate sequence"  
BASE COUNT 78 a 85 c 133 g 96 t 337 others  
ORIGIN  
Query Match 4.3%; Score 15; DB 6; Length 729;  
Best Local Similarity 100.0%; Pred. No. 5.6e-19;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 257 AYTNCNGARGARG 271  
|||||  
Db 482 AYTNCNGARGARG 496  
RESULT 8  
AX068381  
LOCUS AX068381 987 bp DNA linear PAT 25-JAN-2001  
DEFINITION Sequence 4 from Patent WO0102565.  
ACCESSION AX068381  
VERSION AX068381.1 GI:12578542  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct  
artificial sequences.  
REFERENCE 1 (bases 1 to 987)

AUTHORS Holloway,J.L. and Lok,S.  
TITLE Secreted protein zacrp4  
JOURNAL Patent: WO 0102565-A 4 11-JAN-2001;  
ZymoGenetics, Inc. (US)  
FEATURES  
source  
1. .987  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="degenerate nucleotide sequence encoding the  
polypeptide of SEQ ID NO:2"  
variation 1. .987  
/note="Each N is A, T, G or C"  
BASE COUNT 111 a 127 c 204 g 120 t 425 others  
ORIGIN  
Query Match 4.3%; Score 15; DB 6; Length 987;  
Best Local Similarity 100.0%; Pred. No. 4.9e-19;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 337 GCNTGYTGGCNYTN 351  
|||||  
Db 37 GCNTGYTGGCNYTN 51  
RESULT 9  
AR108644  
LOCUS AR108644 1155 bp DNA linear PAT 14-FEB-2001  
DEFINITION Sequence 3 from patent US 6111075.  
ACCESSION AR108644  
VERSION AR108644.1 GI:12824131  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1155)  
AUTHORS Xu,W.-f., Presnell,S.R., Yee,D.P. and Foster,D.C.  
TITLE Protease-activated receptor PAR4 (ZCHEMR2)  
JOURNAL Patent: US 6111075-A 3 29-AUG-2000;  
FEATURES  
source  
1. .1155  
/organism="unknown"  
BASE COUNT 100 a 143 c 217 g 161 t 534 others  
ORIGIN  
Query Match 4.3%; Score 15; DB 6; Length 1155;  
Best Local Similarity 100.0%; Pred. No. 4.6e-19;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 145 TAYWSNGAYCCNWSN 159  
|||||  
Db 919 TAYWSNGAYCCNWSN 933  
RESULT 10  
AX179624  
LOCUS AX179624 2019 bp DNA linear PAT 06-AUG-2001  
DEFINITION Sequence 3 from Patent WO0146418.  
ACCESSION AX179624  
VERSION AX179624.1 GI:15132050  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct  
artificial sequences.  
REFERENCE 1 (bases 1 to 2019)  
AUTHORS Holloway,J.L. and Chandrasekher,Y.A.  
TITLE Human slit polypeptide zslit3  
JOURNAL Patent: WO 0146418-A 3 28-JUN-2001;  
ZymoGenetics, Inc. (US)  
FEATURES  
source  
1. .2019  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="degenerate polynucleotide sequence for zslit3"

```
BASE COUNT      231 a   297 c   354 g   230 t   907 others
ORIGIN

Query Match      4.3%; Score 15; DB 6; Length 2019;
Best Local Similarity 100.0%; Pred. No. 3.6e-19;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 310 WSNYTNMGNGTNGGN 324
|||||
Db 1423 WSNYTNMGNGTNGGN 1437

RESULT 11
LOCUS      AR004980      2277 bp      DNA      linear      PAT 04-DEC-1998
DEFINITION Sequence 2 from patent US 5747317.
ACCESSION  AR004980
VERSION     AR004980.1 GI:3965859
SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE   1 (bases 1 to 2277)
AUTHORS    Cao,Z.
TITLE      Human telomerase RNA interacting protein gene
JOURNAL    Patent: US 5747317-A 2 05-MAY-1998;
FEATURES   Location/Qualifiers
            source
            1..2277
            /organism="unknown"

BASE COUNT      511 a   212 c   395 g   216 t   943 others
ORIGIN

Query Match      4.3%; Score 15; DB 6; Length 2277;
Best Local Similarity 100.0%; Pred. No. 3.4e-19;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 294 YCARTGGAARCAARGA 308
|||||
Db 2106 YCARTGGAARCAARGA 2120

RESULT 12
LOCUS      I26713      20 bp      DNA      linear      PAT 07-OCT-1996
DEFINITION Sequence 14 from patent US 5559220.
ACCESSION  I26713
VERSION     I26713.1 GI:1606583
SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE   1 (bases 1 to 20)
AUTHORS    Roessler,P.G. and Ohlrogge,J.B.
TITLE      Gene encoding acetyl-coenzyme A carboxylase
JOURNAL    Patent: US 5559220-A 14 24-SEP-1996;
FEATURES   Location/Qualifiers
            source
            1..20
            /organism="unknown"

BASE COUNT      1 a   5 c   5 t   6 others
ORIGIN

Query Match      4.0%; Score 14; DB 6; Length 20;
Best Local Similarity 100.0%; Pred. No. 9.1e-16;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 22 CAYGGAAYGCNGT 35
|||||
Db 14 CAYGGAAYGCNGT 1

RESULT 13
LOCUS      AX350993      1026 bp      DNA      linear      PAT 06-FEB-2002
DEFINITION Sequence 3 from patent WO0144281.
ACCESSION  AX167244
VERSION     AX167244.1 GI:14596712
SOURCE      Synthetic construct.
ORGANISM    Synthetic construct.
REFERENCE   1 (bases 1 to 1584)
AUTHORS    Holloway,J.L.

DEFINITION Sequence 27 from Patent WO0190358.
ACCESSION  AX350993
VERSION     AX350993.1 GI:18616369
KEYWORDS   unidentified.
SOURCE     unidentified
           unclassified.
REFERENCE  1
AUTHORS    Gorman,D.M.
TITLE      Mammalian receptor proteins; related reagents and methods
JOURNAL    Patent: WO 0190358-A 27 29-NOV-2001;
           SCHERING CORPORATION (US)
FEATURES   Location/Qualifiers
            source
            1..1026
            /organism="unidentified"
            /db_xref="taxon:32644"
            /note="rodent; surmised Mus musculus"

BASE COUNT      170 a   157 c   156 g   122 t   421 others
ORIGIN

Query Match      4.0%; Score 14; DB 6; Length 1026;
Best Local Similarity 100.0%; Pred. No. 1.6e-16;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 257 AYYTNCCNGARGAR 270
|||||
Db 515 AYYTNCCNGARGAR 528

RESULT 14
LOCUS      AX172419      1134 bp      DNA      linear      PAT 03-JUL-2001
DEFINITION Sequence 3 from Patent WO0144479.
ACCESSION  AX172419
VERSION     AX172419.1 GI:14597534
KEYWORDS   synthetic construct.
SOURCE     synthetic construct
           artificial sequences.
ORGANISM    synthetic construct.
REFERENCE   1 (bases 1 to 1134)
AUTHORS    Conklin,D.C., Yamamoto,G., Cooper,E. and Jaspers,S.R.
TITLE      Galactosyltransferase homolog, znssp8
JOURNAL    Patent: WO 0144479-A 3 21-JUN-2001;
           ZymoGenetics, Inc. (US)
FEATURES   Location/Qualifiers
            source
            1..1134
            /organism="synthetic construct"
            /db_xref="taxon:32630"
            /note="degenerate sequence"

BASE COUNT      230 a   102 c   169 g   174 t   459 others
ORIGIN

Query Match      4.0%; Score 14; DB 6; Length 1134;
Best Local Similarity 100.0%; Pred. No. 1.6e-16;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 326 ARGNTAYGARGCN 339
|||||
Db 800 ARGNTAYGARGCN 813

RESULT 15
LOCUS      AX167244      1584 bp      DNA      linear      PAT 03-JUL-2001
DEFINITION Sequence 3 from Patent WO0144281.
ACCESSION  AX167244
VERSION     AX167244.1 GI:14596712
SOURCE      Synthetic construct.
ORGANISM    Synthetic construct.
REFERENCE   1 (bases 1 to 1584)
AUTHORS    Holloway,J.L.
```

TITLE Human secretin-like g-protein coupled receptor  
 JOURNAL Patent: WO 0144281-A 3 21-JUN-2001;  
 ZymoGenetics, Inc. (US)  
 FEATURES Location/Qualifiers  
 source 1..1584  
 /organism="synthetic construct"  
 /db\_xref="taxon:32630"  
 /note="This degenerate nucleotide sequence encodes the  
 amino acid sequence of SEQ ID NO:2."  
 BASE COUNT 210 a 157 c 245 g 257 t 715 others  
 ORIGIN  
 Query Match 4.0%; Score 14; DB 6; Length 1584;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-16;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 56 GNCNMGNGARYTN 69  
 |  
 Db 374 GNCNMGNGARYTN 387

Search completed: November 7, 2002, 04:21:28  
 Job time : 2775 secs

Barrett  
82 09 649 66 28  
USD-849-050

THIS PAGE BLANK (USPTO)

KW bone marrow transplants; proliferation-restricted cells;  
 XX ambiguity-maximised; ds.

OS Homo sapiens.  
 XX Synthetic.

FH Key Location/Qualifiers  
 FT CDS 1..2277  
 FT /\*tag= a  
 FT /product= telomerase protein p105

XX WO9801543-A1.

XX 15-JAN-1998.

PD 08-JUL-1997; 97WO-US12297.

PF 08-JUL-1996; 96US-0676967.

XX (TULA-) TULARIK INC.

XX Cao Z;

XX WPI; 1998-101044/09.

DR P-PSDB; AAW41927.

XX New nucleic acid encoding human telomerase protein p105 or its  
 PT fragments - used for therapeutic modulation of telomerase activity  
 PT and for screening for potential modulators of telomerase-target  
 PT binding

XX Disclosure; Pages 20-21; 32pp; English.

XX The sequence is that of an ambiguity-maximised human telomerase  
 CC protein coding sequence. The sequence, or specific fragments  
 CC of it, can be used to modulate expression of a telomerase  
 CC transcript (by hybridising to it intracellularly), e.g.  
 CC for treatment or prevention of cancer, restenosis, inflammation,  
 CC myocardial infarction, glomerulonephritis, transplant rejection and  
 CC infections (e.g. with human immunodeficiency virus). It can be used to  
 CC express recombinant telomerase protein which can be used to screen  
 CC for agents, e.g. antibodies, that modulate binding of human telomerase  
 CC to its binding target. Those that inhibit telomerase activity can be  
 CC used to treat the conditions listed above, while those that are  
 CC agonists can be used to extend the life of proliferation-restricted  
 CC cells, especially normal somatic cells, e.g. in cases of  
 CC hypersensitivity or atrophy, also to improve production of  
 CC recombinant proteins by maximising cell density and survival and  
 CC expansion of precursor cells being used for bone marrow transplants.  
 CC They may also be used for diagnosis. Other uses of telomerase proteins  
 CC are isolation, enrichment and concentration of telomerase RNA or  
 CC proteins; as immunogens; in therapy; as reagent where nascent  
 CC oligonucleotides of known structure are needed (e.g. for tagging  
 CC native nucleic acid molecules) and for regulating cell growth/density  
 CC tolerance. The agents and the telomerase proteins should be very  
 CC specific, e.g. they are selective for cancer cells without harming  
 CC somatic cells.

XX Sequence 2277 BP; 513 A; 212 C; 395 G; 217 T; 940 other;

Query Match 4.3%; Score 15; DB 19; Length 2277;  
 Best Local Similarity 100.0%; Pred. No. 0.003;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 294 YCARTGGAARCARGA 308

Db 2106 YCARTGGAARCARGA 2120

RESULT 15

AAV05370

ID AAV05370 standard; cDNA: 2277 BP.

XX

AC AAV05370;

XX 06-JUL-1998 (first entry)

DE Human telomerase p105 subunit synthetic gene.

XX Telomerase; p105; human; cell replication; cancer; restenosis;  
 KW multiple sclerosis; inflammation; rheumatoid arthritis;  
 KW myocardial infarction; glomerulonephritis; transplant rejection;  
 KW infection; therapy; ds.

XX Homo sapiens.

OS Synthetic.

XX WO9801542-A1.

XX 15-JAN-1998.

XX 08-JUL-1997; 97WO-US12296.

XX 08-JUL-1996; 96US-0676974.

XX (REGC ) UNIV CALIFORNIA.

XX Collins K;

XX WPI; 1998-101043/09.

XX New nucleic acid encoding human telomerase proteins or their  
 PT fragments - useful for therapeutic modulation of telomerase activity  
 PT and for screening for potential modulators of telomerase-target  
 PT binding

XX Disclosure; Page 19-20; 32pp; English.

XX This polynucleotide comprises a synthetic, ambiguity-maximised  
 CC DNA coding for the p105 subunit (see AAW46593) of human telomerase.  
 CC It is based on an isolated cDNA clone (see AAV05369) for p105 and  
 CC encompasses all possible nucleic acids encoding the full-length  
 CC protein. The invention provides methods relating to human  
 CC telomerase and related nucleic acids, including the subunit  
 CC proteins p140, p105, p48 and p43. The proteins may be produced  
 CC recombinantly from transformed host cells or purified from human  
 CC cells. Also included are human telomerase RNA (see AAV05373) and  
 CC functional derivatives (see AAV05374 and AAV16092-93), as well as p105  
 CC synthetic DNA sequences (AAV05370-72). The invention also provides  
 CC isolated telomerase hybridisation probes and primers capable of  
 CC specifically hybridising with the telomerase gene, telomerase-  
 CC specific binding agents such as specific antibodies, and methods of  
 CC making and using the subject compositions in diagnosis (e.g.  
 CC genetic hybridisation screens for telomerase transcripts), therapy  
 CC (e.g. gene therapy to modulate telomerase gene expression) and in  
 CC the biopharmaceutical industry (e.g. reagents for screening  
 CC chemical libraries for lead agents). Modulation of telomerase  
 CC expression can be used for the treatment or prevention of cancer,  
 CC restenosis, inflammation, myocardial infarction, glomerulonephritis,  
 CC transplant rejection or infections (e.g. with HIV).

XX Sequence 2277 BP; 511 A; 212 C; 395 G; 216 T; 943 other;

Query Match 4.3%; Score 15; DB 19; Length 2277;

Best Local Similarity 100.0%; Pred. No. 0.003;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 294 YCARTGGAARCARGA 308

Db 2106 YCARTGGAARCARGA 2120

Search completed: November 7, 2002, 03:35:06

Job time : 307 secs

KW platelet; proliferation; differentiation; mediation;  
 KW inflammatory process; vascular injury; chemotaxis; mitogenesis;  
 KW growth factor; production; degenerate; ds.  
 XX Synthetic.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 1..1155  
 FT /\*tag= a  
 FT /product= "Human protease-activated receptor PAR4"  
 FT /note= "No stop codon given in the specification"  
 XX  
 XX WO9950415-A2.  
 XX  
 XX 07-OCT-1999.  
 XX  
 XX 31-MAR-1999; 99WO-US07100.  
 XX  
 XX 01-APR-1998; 98US-0053866.  
 XX  
 XX (ZYMO ) ZYMOGENETICS INC.  
 XX (UNITW ) UNIV WASHINGTON.  
 XX  
 XX Xu W, Presnell SR, Yee DP, Foster DC;  
 XX  
 XX WPI; 1999-633640/54.  
 XX P-PSDB; AAY50135.  
 XX  
 XX Novel protease activated receptor 4, useful for screening for  
 PT (ant)agonists for promoting the proliferation and/or differentiation of  
 PT platelets and in mediating inflammatory events -  
 XX  
 XX Disclosure; Page 81-82; 85pp; English.  
 XX  
 XX This sequence represents degenerate DNA encoding human protease-activated  
 CC receptor PAR4 (also referred to as ZCHEMR2). Protease-activated receptors  
 CC (PARs) are a subfamily of G protein coupled receptors which are capable  
 CC of mediating cellular signalling in response to proteases (e.g.,  
 CC thrombin). They are characterised by a tethered peptide ligand at the  
 CC extracellular N-terminus that is generated by proteolysis. PAR4 is  
 CC activated by thrombin or trypsin cleavage at Arg47/Gly48, which generates  
 CC a new N-terminus corresponding to the tethered ligand (a hexapeptide).  
 CC Agonists of PAR4 are useful for upregulating cellular or physiological  
 CC responses whereas antagonists are used to downregulate these  
 CC activities. The PAR4 protein is further useful for dissecting the  
 CC effects of thrombin or other activating proteases in the clotting  
 CC pathway from the effects of these proteases at the cellular level.  
 CC Agonists are specifically useful in promoting the proliferation  
 CC and/or differentiation of platelets, in promoting inflammatory events,  
 CC responses to vascular injury, chemotaxis or mitogenesis, and in  
 CC producing growth factors. Antagonists are useful as research reagents  
 CC for characterising sites of ligand-receptor interaction.  
 XX  
 SQ Sequence 1155 BP; 100 A; 143 C; 217 G; 161 T; 534 other;  
 Query Match 4.3%; Score 15; DB 20; Length 1155;  
 Best Local Similarity 100.0%; Pred. No. 0.0038;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 145 TAYWSNGAYCCNWSN 159  
 Db 919 TAYWSNGAYCCNWSN 933  
 |||||  
 RESULT 13  
 ID AAH28471  
 XX AAH28471 standard; DNA; 2019 BP.  
 AC  
 XX AAH28471;  
 XX  
 XX 17-SEP-2001 (first entry)  
 XX

DE Degenerate nucleotide sequence of human slit polypeptide Zslit3.  
 XX  
 KW Slit protein; Zslit3; neurite growth; cellular proliferation;  
 KW immune response; stroke; brain damage; paralysis; Huntington's disease;  
 KW neurodegenerative disease; amyotrophic lateral sclerosis;  
 KW Alzheimer's disease; Parkinson's disease; peripheral neuropathy;  
 KW demyelinating disease; multiple sclerosis; lung organogenesis;  
 KW pulmonary disease; respiration; sclerosis; cystic fibrosis; asthma;  
 KW immunosuppression; autoimmune disease; insulin dependent diabetes;  
 KW rheumatoid arthritis; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200146418-A1.  
 XX  
 XX 28-JUN-2001.  
 XX  
 XX 14-DEC-2000; 2000WO-US34230.  
 XX  
 XX 21-DEC-1999; 99US-0469847.  
 XX  
 XX (ZYMO ) ZYMOGENETICS INC.  
 XX  
 XX Holloway JL, Chandrasekher YA;  
 XX WPI; 2001-441677/47.  
 XX  
 XX Novel human slit polypeptide, ZSLIT3, useful for treating and  
 PT diagnosing cystic fibrosis, insulin dependent diabetes and multiple  
 PT sclerosis -  
 XX  
 XX Claim 3; Page 122; 125pp; English.  
 XX  
 XX The present sequence encodes a human slit protein polypeptide,  
 CC designated Zslit3. Zslit is a neurite growth and development modulator,  
 CC and an cellular proliferation and differentiation and immune response  
 CC modulator. Zslit3 polypeptides and polynucleotides are useful for  
 CC regenerating and directing neurite outgrowths following strokes, brain  
 CC damage caused by head injuries, paralysis caused by spinal injuries,  
 CC and for treating neurodegenerative diseases such as amyotrophic lateral  
 CC sclerosis, Alzheimer's disease, Huntington's disease, Parkinson's  
 CC disease and peripheral neuropathies, or demyelinating diseases  
 CC e.g. multiple sclerosis. They are useful for lung organogenesis and  
 CC repair, and thus useful for diagnosing and treating pulmonary diseases  
 CC such as respiration and circulation, cystic fibrosis and asthma. They  
 CC also act as a mediator of immunosuppression, and thus are useful for  
 CC diagnosing and treating autoimmune diseases such as insulin dependent  
 CC diabetes and rheumatoid arthritis.  
 XX  
 SQ Sequence 2019 BP; 231 A; 297 C; 354 G; 230 T; 907 other;  
 Query Match 4.3%; Score 15; DB 22; Length 2019;  
 Best Local Similarity 100.0%; Pred. No. 0.0031;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 310 WSNYTNMGNGTNGCN 324  
 Db 1423 WSNYTNMGNGTNGCN 1437  
 |||||  
 RESULT 14  
 ID AAV13834  
 XX AAV13834 standard; cDNA; 2277 BP.  
 AC  
 XX AAV13834;  
 XX  
 XX 21-JUL-1998 (first entry)  
 XX  
 XX Homo sapiens ambiguity-maximised telomerase protein p105 gene.  
 DE telomerase; p105; treatment; prevention; cancer; restenosis;  
 KW inflammation; myocardial infarction; glomerulonephritis; transplant;  
 KW rejection; infection; HIV; human immunodeficiency virus;

QY 257 AYTNCNGARGARG 271  
Db 482 AYTNCNGARGARG 496  
|||||

RESULT 10  
ABK15336  
ID ABK15336 standard; DNA; 729 BP.

XX AC ABK15336;

XX 08-MAY-2002 (first entry)

XX Human helical protein zalpha51 degenerate coding sequence #2.

XX Human; helical protein; zalpha51; gene; cytokine; muscular;  
KW neuroprotective; gene therapy; vaccine; protein therapy;  
KW neuromuscular disorder; locomotion disorder; ds.

XX OS Homo sapiens.

XX US2002009775-A1.

XX 24-JAN-2002.

XX 16-MAR-2001; 2001US-0810052.

XX 17-MAR-2000; 2000US-190410P.

XX 25-APR-2000; 2000US-199443P.

XX (CONK/) CONKLIN D C.

XX (PRES/) PRESNELL S R.

XX Conklin DC, Presnell SR;

XX WPI; 2002-171143/22.

XX Helical protein zalpha51 and an isolated Glu-Glu affinity tag

XX polypeptide, useful in the prevention, diagnosis and treatment of

XX neuromuscular disorders, e.g. locomotion disorders -

XX Claim 28; Page 31; 4lpp; English.

XX The present invention relates to a new helical protein zalpha51  
CC (comprising a 4 helix bundle) and an isolated Glu-Glu affinity tag  
CC polypeptide, and the nucleic acids encoding them, as defined in the  
CC specification. The nucleic acids and proteins may be used in the  
CC prevention, diagnosis and treatment of diseases associated with  
CC inappropriate zalpha51 expression. The nucleic acids and their  
CC complements may also be used as DNA probes in diagnostic assays to  
CC detect and quantitate the presence of similar nucleic acids in samples,  
CC and therefore which patients may be in need of restorative therapy. The  
CC zalpha51 polypeptide may also be used as antigens in the production of  
CC antibodies against zalpha51 and in assays to identify modulators of  
CC zalpha51 expression and activity. The anti-zalpha51 antibodies and  
CC antagonists may also be used to down regulate expression and activity.  
CC The anti-zalpha51 antibodies may also be used as diagnostic agents for  
CC detecting the presence of zalpha51 polypeptide in samples. Disorders that  
CC may be prevented, diagnosed and/or treated by the above methods include,  
CC for example, neuromuscular disorders, especially a locomotion disorder.  
CC The present nucleic acid sequence represents all nucleic acids that  
CC encode zalpha51 #2.

XX Sequence 729 BP; 78 A; 85 C; 133 G; 96 T; 337 other;

Query Match 4.3%; Score 15; DB 24; Length 729;  
Best Local Similarity 100.0%; Pred. No. 0.0045;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 257 AYTNCNGARGARG 271

Db 482 AYTNCNGARGARG 496

RESULT 11

AAF28673  
ID AAF28673 standard; CDNA; 987 BP.

XX AC AAF28673;

XX 05-APR-2001 (first entry)

XX Degenerate human zacrpa4 coding sequence.

XX Human; zacrpa4; complement factor C1q domain; chromosome 11q11;  
KW energy balance; cellular metabolic reaction; autocrine factor;  
KW development; cell proliferation; differentiation; cell survival; ss.

XX OS Homo sapiens.

XX WO200102565-A2.

XX 11-JAN-2001.

XX 28-JUN-2000; 2000WO-US17692.

XX 01-JUL-1999; 99US-0346502.

XX (ZYMO ) ZYMOGENETICS INC.

XX Holloway JL, Lok S;

XX WPI; 2001-138140/14.

XX P-PSDB; AAB61608.

XX Novel secreted protein ZACRP4 polypeptides having tandem C1q globular  
PT domains, useful for studying cell-cell communication and regulation of  
PT cellular processes -

XX Disclosure; Page 80; 82pp; English.

XX The present sequence is a degenerate coding sequence for human ZACRP4  
CC protein. ZACRP4 protein has two complement factor C1q domains. The ZACRP4  
CC gene is located on human chromosome 11q11. The ZACRP4 coding sequence and  
CC protein have a number of uses described in the specification, including,  
CC modulation of energy balance and cellular metabolic reactions in  
CC mammals. In addition, ZACRP4 protein is useful as an autocrine factor,  
CC particularly during development, in mediating the processes of an  
CC organism, in regulating cellular processes such as cell proliferation  
CC and/or differentiation, cell survival and energy balance.

XX Sequence 987 BP; 111 A; 127 C; 204 G; 120 T; 425 other;

Query Match 4.3%; Score 15; DB 22; Length 987;

Best Local Similarity 100.0%; Pred. No. 0.0041;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 337 GCNTGTYGGGCGNYTN 351

Db 37 GCNTGTYGGGCGNYTN 51

RESULT 12

AAZ32748

ID AAZ32748 standard; DNA; 1155 BP.

XX AC AAZ32748;

XX 31-JAN-2000 (first entry)

XX Human protease-activated receptor PAR4 degenerate DNA.

XX Protease-activated receptor; PAR4; ZCHEMR2; G protein coupled;

XX cellular signalling; protease; tethered ligand; N-terminal;

XX proteolysis; thrombin; trypsin; cleavage; hexapeptide; agonist;

XX antagonist; cellular response; physiological response; clotting pathway;

CC degenerate DNA #1 encompassing all DNAs encoding human zalpha51.  
 XX Sequence 696 BP; 74 A; 82 C; 121 G; 93 T; 326 other;  
 SQ

Query Match 4.3%; Score 15; DB 22; Length 696;  
 Best Local Similarity 100.0%; Pred. No. 0.0046;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 257 AYYTNCNGARGARG 271  
 DB 449 AYYTNCNGARGARG 463  
 |||||

RESULT 8  
 ID ABK15334  
 AC ABK15334 standard; DNA; 696 BP.  
 XX  
 XX  
 XX

DT 08-MAY-2002 (first entry)  
 DE Human helical protein zalpha51 degenerate coding sequence #1.  
 XX

Human; helical protein; zalpha51; gene; cytokine; muscular;  
 KW neuroprotective; gene therapy; vaccine; protein therapy;  
 KW neuromuscular disorder; locomotion disorder; ds.  
 XX  
 XX  
 OS Homo sapiens.  
 XX

PN US2002009775-A1.  
 XX

PD 24-JAN-2002.  
 XX

PF 16-MAR-2001; 2001US-0810052.  
 XX

PR 17-MAR-2000; 2000US-190410P.  
 PR 25-APR-2000; 2000US-199443P.  
 XX

PA (CONK/) CONKLIN D C.  
 PA (PRES/) PRESNELL S R.  
 XX

PI Conklin DC, Presnell SR;  
 XX

DR WPI; 2002-171143/22.  
 XX

PT Helical protein zalpha51 and an isolated Glu-Glu affinity tag  
 PT polypeptide, useful in the prevention, diagnosis and treatment of  
 PT neuromuscular disorders, e.g. locomotion disorders -  
 XX

PS Disclosure; Page 28-29; 41pp; English.  
 XX

CC The present invention relates to a new helical protein zalpha51  
 CC (comprising a 4 helix bundle) and an isolated Glu-Glu affinity tag  
 CC polypeptide, and the nucleic acids encoding them, as defined in the  
 CC specification. The nucleic acids and proteins may be used in the  
 CC prevention, diagnosis and treatment of diseases associated with  
 CC inappropriate zalpha51 expression. The nucleic acids and their  
 CC complements may also be used as DNA probes in diagnostic assays to  
 CC detect and quantitate the presence of similar nucleic acids in samples,  
 CC and therefore which patients may be in need of restorative therapy. The  
 CC zalpha51 polypeptide may also be used as antigens in the production of  
 CC antibodies against zalpha51 and in assays to identify modulators of  
 CC zalpha51 expression and activity. The anti-zalpha51 antibodies and  
 CC antagonists may also be used to down regulate expression and activity.  
 CC The anti-zalpha51 antibodies may also be used as diagnostic agents for  
 CC detecting the presence of zalpha51 polypeptide in samples. Disorders that  
 CC may be prevented, diagnosed and/or treated by the above methods include,  
 CC for example, neuromuscular disorders, especially a locomotion disorder.  
 CC The present nucleic acid sequence represents all nucleic acids that  
 CC encode zalpha51 #1.  
 XX

SQ Sequence 696 BP; 74 A; 82 C; 121 G; 93 T; 326 other;

Query Match 4.3%; Score 15; DB 24; Length 696;  
 Best Local Similarity 100.0%; Pred. No. 0.0046;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 257 AYYTNCNGARGARG 271  
 DB 449 AYYTNCNGARGARG 463  
 |||||

RESULT 9  
 ID AAS13926 standard; DNA; 729 BP.  
 XX  
 AC AAS13926;  
 XX  
 XX

DT 19-DEC-2001 (first entry)  
 DE Degenerate DNA #2 encompassing all DNAs encoding human zalpha51.  
 XX

Human; zalpha51; antirheumatic; antiarthritic; neuroprotective;  
 KW antidiabetic; antinflammatory; dermatological; immunosuppressive;  
 KW antidiabetic; antibacterial; cytostatic; neuromuscular disorder;  
 KW locomotion disorder; cancer; tumour imaging; immune disorder;  
 KW graft rejection; graft-versus-host disease; ds.  
 XX  
 OS Synthetic.  
 XX

PN WO200170986-A2.  
 XX

PD 27-SEP-2001.  
 XX

PF 16-MAR-2001; 2001WO-US08493.  
 XX

PR 17-MAR-2000; 2000US-0527843.  
 PR 25-APR-2000; 2000US-0558459.  
 XX

PA (ZYMO ) ZYMOGENETICS INC.  
 XX

PI Conklin DC, Presnell SR;  
 XX

DR WPI; 2001-639129/73.  
 XX

PT Novel cytokine, zalpha51, useful as diagnostic for liver,  
 PT neuroblastoma, brain and other cancers, and for treating rheumatoid  
 PT arthritis, multiple sclerosis, myasthenia gravis, systemic lupus  
 PT erythematosis, diabetes -  
 XX

PS Claim 28; p89; 89pp; English.  
 XX

CC The invention relates to a novel isolated zalpha 51 polypeptide (a  
 CC cytokine) (I). The nucleic acid (II) encoding zalpha51 is useful as  
 CC a probe for detecting the presence of RNA encoding (I) in a biological  
 CC sample. (II) is also useful as a diagnostic reagent for detecting a  
 CC genetic abnormality in a patient. The antibody to (I) is useful for  
 CC detecting the presence of (I) in a biological sample taken from a  
 CC mammal with a neuromuscular disorder or locomotion disorder. (I) is  
 CC useful for raising antibodies, including monoclonal antibodies  
 CC that specifically bind to (I). The zalpha51 antagonist is useful for  
 CC treating cancer, and also for modulating the immune system, reducing  
 CC graft rejection, preventing graft-versus-host disease, boosting  
 CC immunity to infectious diseases, treating immuno-compromised patients  
 CC or improving vaccines. (I), (II), and the antibody to (I) are useful  
 CC for diagnosing and treating disorders associated with abnormal cell  
 CC proliferation. Labeled zalpha51 polypeptides may also be used for  
 CC imaging tumours. The present sequence represents the coding sequence of  
 CC degenerate DNA #2 encompassing all DNAs encoding human zalpha51.  
 XX

SQ Sequence 729 BP; 78 A; 85 C; 133 G; 96 T; 337 other;

Query Match 4.3%; Score 15; DB 22; Length 729;  
 Best Local Similarity 100.0%; Pred. No. 0.0045;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



CC	cardiac function, induce skeletal muscle neogenesis and/or hyperplasia,
CC	be used in kidney regeneration and/or for treating systemic and
CC	pulmonary hypertension. Zalpha31 can be used for treating viral
CC	leukaemias, acquired immunodeficiency syndrome (AIDS) or gastrointestinal
CC	viral infections.
XX	
SQ	Sequence 426 BP; 101 A; 41 C; 70 G; 43 T; 171 other;
	Query Match 4.3%; Score 15; DB 22; Length 426;
	Best Local Similarity 100.0%; Pred. No. 0.0055;
	Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	263 CNGARGAGCNAARG 277
Db	77 CNGARGAGCNAARG 91
RESULT 7	
AAS13924	
ID	AAS13924 standard; DNA; 696 BP.
XX	
AC	AAS13924;
XX	
DT	19-DEC-2001 (first entry)
XX	
DE	Degenerate DNA #1 encompassing DNAs encoding human Zalpha51.
XX	
KW	Human; Zalpha51; antirheumatic; antiarthritic; neuroprotective;
KW	antidiabetic; antiinflammatory; dermatological; immunosuppressive;
KW	antisthmatic; antibacterial; cytostatic; neuromuscular disorder;
KW	locomotion disorder; cancer; tumour imaging; immune disorder;
KW	graft rejection; graft-versus-host disease; ds.
XX	
OS	Synthetic.
XX	
PN	WO200170986-A2.
XX	
PD	27-SEP-2001.
XX	
PF	16-MAR-2001; 2001WO-US08493.
XX	
PR	17-MAR-2000; 2000US-0527843.
PR	25-APR-2000; 2000US-0558459.
XX	
PA	(ZYM0 ) ZYMOGENETICS INC.
XX	
PI	Conklin DC, Presnell SR;
XX	
DR	WPI; 2001-639129/73.
XX	
PT	Novel cytokine, Zalpha51, useful as diagnostic for liver,
PT	neuroblastoma, brain and other cancers, and for treating rheumatoid
PT	arthritis, multiple sclerosis, myasthenia gravis, systemic lupus
PT	erythematosus, diabetes
XX	
PS	Disclosure; p85; 89pp; English.
XX	
CC	The invention relates to a novel isolated Zalpha 51 polypeptide (a
CC	cytokine) (I). The nucleic acid (II) encoding Zalpha51 is useful as
CC	a probe for detecting the presence of RNA encoding (I) in a biological
CC	sample. (II) is also useful as a diagnostic reagent for detecting a
CC	genetic abnormality in a patient. The antibody to (I) is useful for
CC	detecting the presence of (I) in a biological sample taken from a
CC	mammal with a neuromuscular disorder or locomotion disorder. (I) is
CC	useful for raising antibodies, including monoclonal antibodies
CC	that specifically bind to (I). The Zalpha51 antagonist is useful for
CC	treating cancer, and also for modulating the immune system, reducing
CC	graft rejection, preventing graft-versus-host disease, boosting
CC	immunity to infectious diseases, treating immuno-compromised patients
CC	or improving vaccines. (I), (II), and the antibody to (I) are useful
CC	for diagnosing and treating disorders associated with abnormal cell
CC	proliferation. Labeled Zalpha51 polypeptides may also be used for
CC	imaging tumours. The present sequence represents the coding sequence of

CC OX2RH1, OX2RH2, OX2RH3, OX2RH4, or OX2RH1.2 proteins. Agonists and  
CC antagonists of the OX2RH sequences can be used to modulate physiology  
CC or development of a cell, particularly for enhancing myeloid function or  
CC enhancing immunity. The sequences can be used to identify non-OX2 ligands  
CC for an OX2R. The polypeptides and polynucleotides can be used to treat  
CC inflammatory, leukoproliferative, neurodegenerative or post-traumatic  
CC conditions, including atherosclerosis, multiple sclerosis, ischaemia,  
CC neurodegeneration, rheumatoid arthritis, and autoimmunity. The present  
CC sequence represents the reverse translated degenerate polynucleotide  
CC sequence of human OX2R homologue 1 (OX2RH1).

XX SQ Sequence 885 BP; 182 A; 98 C; 133 G; 114 T; 358 other;  
Query Match 4.6%; Score 16; DB 22; Length 885;  
Best Local Similarity 100.0%; Pred. No. 0.00061; Mismatches 0; Indels 0; Gaps 0;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 AAYGTNWSNACNGTNA 244  
|||||  
Db 697 AAYGTNWSNACNGTNA 712

RESULT 4  
AAC84178  
ID AAC84178 standard; DNA: 1044 BP.  
XX AAC84178;  
AC AAC84178;  
XX 19-MAR-2001 (first entry)  
DT 19-MAR-2001 (first entry)  
XX Human OX2RH1.2 degenerate nucleotide sequence.  
DE Human OX2RH1.2 degenerate nucleotide sequence.  
XX OX2R protein: OX2RH1; OX2RH2; OX2RH3; OX2RH4; OX2RH1.2; ischaemia;  
KW antiinflammatory; cytostatic; neuroprotective; nootropic; human;  
KW antiarteriosclerotic; vasotropic; immunosuppressive; antirheumatic;  
KW antiarthritic; gene therapy; ds.  
XX  
OS Homo sapiens.  
XX  
XX W0200070045-A1.  
PN W0200070045-A1.  
XX  
XX 23-NOV-2000.  
XX  
XX 11-MAY-2000; 2000WO-US12998.  
XX  
XX 13-MAY-1999; 99GB-0011123.  
PR 03-NOV-1999; 99GB-0025989.  
XX  
XX (MEDI-) MEDICAL RES COUNCIL.  
PA (SCHE) SCHERING CORP.  
XX  
XX Barclay AN, Brown MH, Gorman DM, Lanier LL, Wright GJ;  
PI Cherwinski H, Phillips JH, Hoek RM, Sedgwick JD;  
XX  
XX WPI; 2001-016233/02.  
DR P-PSDB; AAB48016.  
XX  
XX Mammalian OX2R proteins and DNA sequences useful for modulating the  
PT physiology and development of a cell -  
PT  
XX Disclosure; Page 36; 142pp; English.  
XX  
XX The invention relates to rodent or primate OX2R proteins, especially  
CC OX2RH1, OX2RH2, OX2RH3, OX2RH4, or OX2RH1.2 proteins. Agonists and  
CC antagonists of the OX2RH sequences can be used to modulate physiology  
CC or development of a cell, particularly for enhancing myeloid function or  
CC enhancing immunity. The sequences can be used to identify non-OX2 ligands  
CC for an OX2R. The polypeptides and polynucleotides can be used to treat  
CC inflammatory, leukoproliferative, neurodegenerative or post-traumatic  
CC conditions, including atherosclerosis, multiple sclerosis, ischaemia,  
CC neurodegeneration, rheumatoid arthritis, and autoimmunity. The present  
CC sequence represents the reverse translated degenerate polynucleotide  
CC sequence of human OX2R homologue 1.2 (OX2RH1.2).

XX SQ Sequence 1044 BP; 220 A; 117 C; 154 G; 131 T; 422 other;  
Query Match 4.6%; Score 16; DB 22; Length 1044;  
Best Local Similarity 100.0%; Pred. No. 0.00058;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 AAYGTNWSNACNGTNA 244  
|||||  
Db 697 AAYGTNWSNACNGTNA 712

RESULT 5  
AAH74766  
ID AAH74766 standard; DNA: 2379 BP.  
XX AAH74766;  
AC AAH74766;  
XX 29-OCT-2001 (first entry)  
DT 29-OCT-2001 (first entry)  
XX Degenerate nucleotide sequence of a human zdint5 polypeptide.  
DE Degenerate nucleotide sequence of a human zdint5 polypeptide.  
XX Human; zdint5; anti-angiogenic; intestinal polypeptide; wound healing;  
KW extracellular matrix interaction; tumour suppression; gamete maturation;  
KW immunologic recognition; gastrointestinal irradiation; chemotherapy;  
KW proteolysis; apoptosis; angiogenesis; infection; cell adhesion;  
KW cell fusion; cell signalling; tumour; Crohn's disease; melanoma; trauma;  
KW inflammatory bowel disease; food poisoning; degenerative disease;  
KW inflammation; fertility; gamete maturation; epithelial disorder; ss.  
XX  
XX Synthetic.  
OS Homo sapiens.  
XX  
XX W0200159112-A1.  
PN W0200159112-A1.  
XX  
XX 16-AUG-2001.  
XX  
XX 09-FEB-2001; 2001WO-US04198.  
PF  
XX 10-FEB-2000; 2000US-0501806.  
PR  
XX (ZYMO) ZYMOGENETICS INC.  
XX  
XX Holloway JL, Sheppard PO;  
PI  
XX WPI; 2001-522477/57.  
XX  
XX New anti-angiogenic intestinal polypeptides, zdint5 polypeptides, which  
PT are members of disintegrin proteases, for modulating extracellular  
PT matrix interaction, tumour suppression and wound healing -  
XX  
XX Disclosure; Page 88-89; 92pp; English.  
XX  
XX The present sequence encodes a human zdint5 polypeptide. The zdint5  
CC polypeptide is an anti-angiogenic intestinal polypeptide. Zdint5 is  
CC used for modulating extracellular matrix interactions. Zdint5  
CC polypeptide is useful as a tool for identifying new family members  
CC of polypeptides. Zdint5 polynucleotides are useful as probes or primers  
CC to clone 5' non-coding regions of zdint5 gene. Zdint5 polypeptides are  
CC used for tumour suppression, gamete maturation, immunologic recognition,  
CC and growth and differentiation either working in isolation or in  
CC conjunction with other molecules in colon, small intestine, fetal lung,  
CC testis and B-cells. zdint5 polypeptides are also useful for promoting  
CC wound healing, in the treatment of disorders associated with recovery  
CC after gastrointestinal irradiation, chemotherapy or antibody use,  
CC as anti-infectives, and extracellular matrix repair and remodeling. The  
CC polypeptides are also useful for modulating proteolysis, apoptosis,  
CC angiogenesis, infection, cell adhesion, cell fusion and signalling.  
CC The polypeptides are also useful for treating tumour formation, Crohn's  
CC disease, inflammatory bowel disease, food poisoning, melanoma,  
CC degenerative diseases, disorders related to immunity, inflammation,  
CC fertility, gamete maturation, immunology, trauma and epithelial  
CC disorders.

PT neurodegenerative disorders or cancers  
 XX  
 PS Example 1; Fig 2; 87pp; English.  
 XX  
 CC Primer ROS1 is based on an amino acid sequence of the receptor  
 CC tyrosine kinase (RTK) ROS. A series of primers (see AAV42003-11)  
 CC based on unique motifs of LMR1\_r (see AAW48841) or on unrelated  
 CC RTKs TEK and ROS were used in the PCR amplification of cDNA from  
 CC rat, human and mouse sources. cDNA clones (see AAV32448-56) coding  
 CC for novel RTKs designated LMR1, LMR2 and LMR3 (see AAW48841-49)  
 CC were obtained. LMRs belong to a unique family of receptors and  
 CC are expressed in neuronal tissues and tumour cells, making them  
 CC targets for the treatment of neurodegenerative disorders and  
 CC cancer.  
 XX  
 SQ Sequence 23 BP; 3 A; 1 C; 8 G; 4 T; 7 other;  
 Query Match 4.6%; Score 16; DB 19; Length 23;  
 Best Local Similarity 100.0%; Pred. No. 0.0023;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 322 GGGGARGTNTAYGARG 337  
 Db 7 GGGGARGTNTAYGARG 22  
 RESULT 2  
 AA240496  
 ID AA240496 standard; DNA; 23 BP.  
 XX  
 AC AA240496;  
 DT  
 DT 18-FEB-2000 (first entry)  
 XX  
 DE Degenerate primer ROS1 for STE20 related protein kinase genes.  
 XX  
 KW Antirheumatic; antiarthritic; antiinflammatory; antiallergic; osteopathic;  
 KW antipsoriatic; antiarteriosclerotic; antiasthmatic; immunosuppressive;  
 KW neuroprotective; cardiac; cerebroprotective; cytoskeletal; antidiabetic;  
 KW vulnery; STE20; protein kinase; STLK2; STLK3; STLK4; STLK5; STLK6; STLK7;  
 KW ZC1; ZC2; ZC3; ZC4; KHS2; SULU1; SULU3; GEK2; PAK4; PAK5; antagonist;  
 KW antibody; gene therapy; rheumatoid arthritis; artherosclerosis; asthma;  
 KW inflammatory bowel disease; Crohn's disease; osteoarthritis; psoriasis;  
 KW rhinitis; autoimmunity; organ transplantation; multiple sclerosis;  
 KW myocardial infarction; cardiovascular disease; stroke; renal failure;  
 KW oxidative stress-related neurodegenerative disorder; Parkinson's disease;  
 KW amyotrophic lateral sclerosis; Leigh syndrome; cancer; cardiomyopathy;  
 KW ischemic disorder; inflammation; diabetes mellitus; fibrosis; mitosis;  
 KW mesangial disorder; growth regulation; wound healing; T cell activation;  
 KW immunosuppressant; primer; PCR; amplification; ss.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 XX  
 PN WO9953036-A2.  
 XX  
 PD 21-OCT-1999.  
 XX  
 PF 13-APR-1999; 99WO-US08150.  
 XX  
 PR 14-APR-1998; 98US-0081784.  
 XX  
 PA (SUGE-) SUGEN INC.  
 XX  
 PI Plowman G, Martinez R, Whyte D;  
 XX  
 DR WPI; 1999-611301/52.  
 XX  
 DR Novel kinase-related polypeptides used for the diagnosis and treatment  
 PT of kinase-related diseases and disorders -  
 PT  
 XX Disclosure; Page 316; 387pp; English.  
 PS  
 XX

CC This sequence represents a degenerate oligonucleotide used to isolate the  
 CC coding sequence for a novel STE20-related protein kinase. The invention  
 CC relates to nucleic acid molecule encoding a kinase polypeptide selected  
 CC from the kinases STLK2, STLK3, STLK4, STLK5, STLK6, STLK7, ZC1, ZC2, ZC3,  
 CC ZC4, KHS2, SULU1, SULU3, GEK2, PAK4 and PAK5. The proteins are used to  
 CC identify agonists and antagonists, and to raise antibodies. The  
 CC polynucleotides are useful in gene therapy protocols. The polynucleotides,  
 CC polypeptides, antibodies, antagonists and agonists may be used to treat  
 CC diseases such as immune-related disorders and diseases (e.g. rheumatoid  
 CC arthritis, artherosclerosis, chronic inflammatory bowel disease (e.g.  
 CC Crohn's disease), asthma, osteoarthritis, psoriasis, atherosclerosis,  
 CC rhinitis, autoimmunity, and organ transplantation, chronic inflammatory  
 CC pelvic disease, multiple sclerosis, organ transplantation, myocardial  
 CC infarction, cardiovascular disease, stroke, renal failure, oxidative  
 CC stress-related neurodegenerative disorders (e.g. amyotrophic lateral  
 CC sclerosis, Parkinson's disease and Leigh syndrome), cancer,  
 CC cardiomyopathies, ischemic disorders, inflammatory disorders, diabetes  
 CC mellitus, fibrotic and mesangial disorders. The proteins may also be  
 CC useful for cell growth regulation (e.g. in wound healing), T cell  
 CC activation, mitosis control, and as immunosuppressants.  
 XX  
 SQ Sequence 23 BP; 3 A; 1 C; 8 G; 4 T; 7 other;  
 Query Match 4.6%; Score 16; DB 20; Length 23;  
 Best Local Similarity 100.0%; Pred. No. 0.0023;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 322 GGGGARGTNTAYGARG 337  
 Db 7 GGGGARGTNTAYGARG 22  
 RESULT 3  
 AAC84168  
 ID AAC84168 standard; DNA; 885 BP.  
 XX  
 AC AAC84168;  
 XX  
 DT 19-MAR-2001 (first entry)  
 XX  
 DE Human OX2RH1 degenerate nucleotide sequence.  
 XX  
 KW OX2R protein; OX2RH1; OX2RH2; OX2RH3; OX2RH4; OX2RH1.2; ischaemia;  
 KW antiinflammatory; cytoskeletal; neuroprotective; nootropic; human;  
 KW antiarteriosclerotic; vasotropic; immunosuppressive; antirheumatic;  
 KW antiarthritic; gene therapy; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200070045-A1.  
 XX  
 PD 23-NOV-2000.  
 XX  
 PF 11-MAY-2000; 2000WO-US12998.  
 XX  
 PR 13-MAY-1999; 99GB-0011123.  
 PR 03-NOV-1999; 99GB-0025989.  
 XX  
 PA (MEDI-) MEDICAL RES COUNCIL.  
 PA (SCHE-) SCHERING CORP.  
 XX  
 PI Barclay AN, Brown MH, Gorman DM, Lanier LL, Wright GJ;  
 PI Cherwinski H, Phillips JH, Hoek RM, Sedgwick JD;  
 XX  
 DR WPI; 2001-016233/02.  
 DR P-PSDB; AAB48011.  
 XX  
 PT Mammalian OX2R proteins and DNA sequences useful for modulating the  
 PT physiology and development of a cell -  
 XX  
 PS Disclosure; Page 35; 142pp; English.  
 XX  
 CC The invention relates to rodent or primate OX2R proteins, especially

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 02:17:05 ; Search time 301 seconds  
(without alignments)  
2626.084 Million cell updates/sec

Title: US09897438BK-2

Perfect score: 351

Sequence: 1 GARCARTGYGCNACNATNAT.....AYGARGCNTGYTGGCNYTN 351

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 112599159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_101002.\*

```

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

```

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	16	4.6	23	AAV42004	Receptor tyrosine
2	16	4.6	23	AAZ40496	Degenerate primer
3	16	4.6	885	AAAC84168	Human OX2RH1 degen
4	16	4.6	1044	AAAC84178	Human OX2RH1.2 deg
5	16	4.6	2379	AAH74766	Degenerate nucleot
6	15	4.3	426	AAAC91387	Human zaipha31 deg
7	15	4.3	696	AAAS13924	Degenerate DNA #1
8	15	4.3	696	ABK15334	Human helical prot
9	15	4.3	729	AAAS13926	Degenerate DNA #2

10	15	4.3	729	24	ABK15336	Human helical prot
11	15	4.3	987	22	AAF28673	Degenerate human z
12	15	4.3	1155	20	AAZ32748	Human protease-act
13	15	4.3	2019	22	AAH28471	Degenerate nucleot
14	15	4.3	2277	19	AAV13834	Homo sapiens ambig
15	15	4.3	2277	19	AAV05370	Human telomerase p
16	14	4.0	20	17	AAAT45730	Primer PR2 used to
17	14	4.0	705	20	AAAX08719	z119a degenerate p
18	14	4.0	1026	24	AAAS18143	Mouse DCRS10 rever
19	14	4.0	1134	22	AAAD09356	Human beta-1,3-gal
20	14	4.0	1218	20	AAAX85600	Degenerate DNA enc
21	14	4.0	1584	22	AAF90613	Secretin-like rece
22	14	4.0	1629	22	AAF90614	Secretin-like rece
23	14	4.0	1695	24	AAAS18141	Human DCRS10 rever
24	14	4.0	1728	22	AAH28626	Human Kunitz domai
25	14	4.0	1728	24	ABK14096	Human zkun8 degene
26	14	4.0	2094	24	AAAD24223	Murine zcytor14 cy
27	14	4.0	2094	24	AAAS18133	Mouse DCRS7 revers
28	13	3.7	17	18	AAAT85341	Spider silk protei
29	13	3.7	20	14	AAQ40096	Thrombin-inhibiti
30	13	3.7	20	16	AAQ404020	Glutathione-indepe
31	13	3.7	23	15	AAQ70814	PCR primer used fo
32	13	3.7	23	16	AAQ92949	Saccharomyces cere
33	13	3.7	51	24	AAAD38244	Human zsig33-gamma
34	13	3.7	51	24	AAAD25765	Human zsig33-gamma
35	13	3.7	351	22	AAAF83679	Human zsig33 degen
36	13	3.7	390	22	AAAD08417	Human DNAX cytokin
37	13	3.7	450	19	AAV45445	Human chemokine zS
38	13	3.7	669	20	AAAX61299	Human 2-19 protein
39	13	3.7	693	22	AAAD06415	Human DNAX cytokin
40	13	3.7	693	22	AAAF83736	Human cytokine rec
41	13	3.7	696	24	AAAD22922	Human soluble IL-2
42	13	3.7	729	24	AAAS18139	Mouse DCRS9 revers
43	13	3.7	747	22	AAAD06411	Human DNAX cytokin
44	13	3.7	783	21	AAZ52250	Human stomach prot
45	13	3.7	804	24	ABK48294	Degenerate DNA enc

#### ALIGNMENTS

##### RESULT 1

AAV42004

ID AAV42004 standard; DNA; 23 BP.

XX AAV42004;

AC AAV42004;

XX 12-OCT-1998 (first entry)

XX Receptor tyrosine kinase ROS-based primer ROS1.

DE Receptor tyrosine kinase; LMR1; LMR2; LMR3; signal transduction;

XX Receptor tyrosine kinase; LMR1; LMR2; LMR3; signal transduction;

KW cancer; cancer; neurodegenerative disorder; gene therapy; PCR;

KW primer; ROS; ss.

XX OS Synthetic.

XX WO9822507-A2.

PN 28-MAY-1998.

XX 21-NOV-1997; 97WO-US22526.

XX 22-NOV-1996; 96US-0031675.

PR (SUGEN-) SUGEN INC.

XX Joho KE, Plowman GD;

XX WPI; 1998-312419/27.

XX New isolated receptor tyrosine kinase genes - which are expressed in

PT neuronal tissues and tumour cells, useful as targets for

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 03:29:56 ; Search time 67 seconds  
(without alignments)  
1606.620 Million cell updates/sec

Title: US09897438BK-2  
Perfect score: 351  
Sequence: 1 GARGCTGGGACNATAT.....AYGARGCTGGTGGCNYTN 351

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 15338381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCRUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	16	4.6	23	3	US-08-976-255-35
2	15	4.3	1155	3	US-09-053-866-3
3	15	4.3	1155	4	US-09-479-130-3
4	15	4.3	2277	1	US-08-676-967-2
5	15	4.3	2277	1	US-08-676-974-2
6	15	4.3	2277	2	US-09-098-487-2
7	14	4.0	20	1	US-08-418-893D-14
8	14	4.0	705	4	US-09-167-513-8
9	13	3.7	20	1	US-08-859-183-3
10	13	3.7	23	1	US-08-447-500-13
11	13	3.7	23	1	US-08-454-097-13
12	13	3.7	23	1	US-08-453-866-13
13	13	3.7	23	3	US-08-185-359-13
14	13	3.7	72	1	US-08-457-192-2
15	13	3.7	1602	4	US-09-124-541-10
16	13	3.7	2082	4	US-09-440-325A-2
17	13	3.7	2265	3	US-09-369-618-3
18	13	3.7	2265	3	US-09-369-617-3
19	12	3.4	19	2	US-08-849-536A-17
20	12	3.4	26	6	5187262-4
21	12	3.4	34	1	US-08-145-006C-10
22	12	3.4	34	5	PCT-US94-00545-10
23	12	3.4	624	4	US-09-397-992A-3
24	12	3.4	624	4	US-09-397-992A-6
25	12	3.4	954	2	US-08-973-275-2
26	12	3.4	1035	4	US-09-457-066-6
27	12	3.4	1176	3	US-09-072-384-16

28	12	3.4	1191	4	US-09-459-133-3	Sequence 3, Appli
29	12	3.4	1659	1	US-08-231-729B-1	Sequence 1, Appli
30	12	3.4	1659	1	US-08-231-729B-2	Sequence 2, Appli
31	12	3.4	2406	4	US-09-632-098-5	Sequence 5, Appli
32	12	3.4	2439	4	US-09-632-098-6	Sequence 6, Appli
33	11	3.1	17	1	US-08-276-452A-45	Sequence 45, Appli
34	11	3.1	17	2	US-08-798-744-45	Sequence 45, Appli
35	11	3.1	18	5	PCT-US94-01321-32	Sequence 32, Appli
36	11	3.1	23	1	US-08-082-844-9	Sequence 9, Appli
37	11	3.1	23	5	PCT-US94-07040-9	Sequence 9, Appli
38	11	3.1	24	1	US-08-275-983B-5	Sequence 5, Appli
39	11	3.1	24	4	US-08-794-002-13	Sequence 13, Appli
40	11	3.1	24	4	US-08-854-039B-13	Sequence 13, Appli
41	11	3.1	27	1	US-08-195-075-11	Sequence 11, Appli
42	11	3.1	27	5	PCT-US95-01672-11	Sequence 11, Appli
43	11	3.1	28	1	US-07-868-353A-24	Sequence 24, Appli
44	11	3.1	28	1	US-08-407-804-33	Sequence 33, Appli
45	11	3.1	28	3	US-09-124-807-33	Sequence 33, Appli

## ALIGNMENTS

RESULT 1  
US-08-976-255-35  
; Sequence 35, Application US/08976255  
; Patent No. 6136581  
; GENERAL INFORMATION:  
; APPLICANT: Jono, Keith E.  
; APPLICANT: Plowman, Gregory  
; TITLE OF INVENTION: KINASE GENES AND USES  
; NUMBER OF SEQUENCES: 53  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90071-2066  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; MEDIUM TYPE: storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM P.C. DOS 5.0  
; SOFTWARE: FASTSEQ for Windows 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/976,255  
; FILING DATE: No. 6136581ember 21, 1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/031,675  
; FILING DATE: No. 6136581ember 22, 1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warburg, Richard J.  
; REGISTRATION NUMBER: 32,327  
; REFERENCE/DOCKET NUMBER: 229/182  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELEX: 67-3510  
; INFORMATION FOR SEQ ID NO: 35:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 23 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; OTHER INFORMATION: "N" stands for A, C, G or T.  
; OTHER INFORMATION: "Y" stands for C or T.  
; OTHER INFORMATION: "R" stands for A or G.  
US-08-976-255-35

Query Match 4.6%; Score 16; DB 3; Length 23;  
Best Local Similarity 100.0%; Pred. No. 0.00046;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 322 GNGARGTNTAYGARG 337  
|||||  
Db 7 GNGARGTNTAYGARG 22

## RESULT 2

US-09-053-866-3  
; Sequence 3, Application US/09053866  
; Patent No. 611075  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Wenfeng  
; APPLICANT: Presnell, Scott R.  
; APPLICANT: Yee, David P.  
; APPLICANT: Foster, Donald C.  
; TITLE OF INVENTION: PROTEASE-ACTIVATED RECEPTOR  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ZymoGenetics, Inc.  
; STREET: 1201 Eastlake Avenue East  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98102  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/053.866  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Leith, Debra K  
; REGISTRATION NUMBER: 32,619  
; REFERENCE/DOCKET NUMBER: 98-10  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206-442-6674  
; TELEFAX: 206-442-6678  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1155 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-053-866-3

Query Match 4.3%; Score 15; DB 3; Length 1155;  
Best Local Similarity 100.0%; Pred. No. 0.00082;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 145 TAYWSNGAYCCNWSN 159  
|||||  
Db 919 TAYWSNGAYCCNWSN 933

## RESULT 3

US-09-479-130-3  
; Sequence 3, Application US/09479130  
; Patent No. 6436400  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Wenfeng  
; APPLICANT: Presnell, Scott R.  
; APPLICANT: Yee, David P.

; APPLICANT: Foster, Donald C.  
; TITLE OF INVENTION: PROTEASE-ACTIVATED RECEPTOR  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ZymoGenetics, Inc.  
; STREET: 1201 Eastlake Avenue East  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98102  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/479.130  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Leith, Debra K  
; REGISTRATION NUMBER: 32,619  
; REFERENCE/DOCKET NUMBER: 98-10  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206-442-6674  
; TELEFAX: 206-442-6678  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1155 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-479-130-3

Query Match 4.3%; Score 15; DB 4; Length 1155;  
Best Local Similarity 100.0%; Pred. No. 0.00082;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 145 TAYWSNGAYCCNWSN 159  
|||||  
Db 919 TAYWSNGAYCCNWSN 933

## RESULT 4

US-08-676-967-2  
; Sequence 2, Application US/08676967  
; Patent No. 5747317  
; GENERAL INFORMATION:  
; APPLICANT: COLLINS, KATHLEEN  
; TITLE OF INVENTION: Human Telomerase  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Science & Technology Law Group  
; STREET: 268 Bush Street, Suite 3200  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/676.967  
; FILING DATE:  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:

NAME: Osman Ph.D., Richard A  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: UCB96-055  
TELEPHONE: (415)343-4341  
TELEFAX: (415)343-4342  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2277 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
US-08-676-967-2

Query Match 4.3%; Score 15; DB 1; Length 2277;  
Best Local Similarity 100.0%; Pred. No. 0.00066;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 294 YCARTGGAARCARGA 308  
|||||

Db 2106 YCARTGGAARCARGA 2120

## RESULT 5

US-08-676-974-2  
Sequence 2, Application US/08676974  
Patent No. 5770422  
GENERAL INFORMATION:  
APPLICANT: COLLINS, KATHLEEN  
TITLE OF INVENTION: Human Telomerase  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Science & Technology Law Group  
STREET: 268 Bush Street, Suite 3200  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94104

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/676,974  
FILING DATE:

## CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:  
NAME: Osman Ph.D., Richard A  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: UCB96-055  
TELEPHONE: (415)343-4341  
TELEFAX: (415)343-4342  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2277 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
US-08-676-974-2

Query Match 4.3%; Score 15; DB 1; Length 2277;  
Best Local Similarity 100.0%; Pred. No. 0.00066;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 294 YCARTGGAARCARGA 308  
|||||

Db 2106 YCARTGGAARCARGA 2120

## RESULT 6

US-09-098-487-2  
Sequence 2, Application US/09098487  
Patent No. 5917025  
GENERAL INFORMATION:  
APPLICANT: COLLINS, Kathleen  
TITLE OF INVENTION: Human Telomerase  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Science & Technology Law Group  
STREET: 268 Bush Street, Suite 3200  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/098,487  
FILING DATE:

## CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:  
NAME: Osman Ph.D., Richard A  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: UCB96-055  
TELEPHONE: (415)343-4341  
TELEFAX: (415)343-4342  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2277 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
US-09-098-487-2

## Query Match

Best Local Similarity 100.0%; Pred. No. 0.00066;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 294 YCARTGGAARCARGA 308

Db 2106 YCARTGGAARCARGA 2120

## RESULT 7

US-08-418-893D-14/c  
Sequence 14, Application US/08418893D  
Patent No. 5559220  
GENERAL INFORMATION:  
APPLICANT: ROESSLER, PAUL G  
APPLICANT: OHLROGE, JOHN B  
TITLE OF INVENTION: GENE THAT ENCODES ACETYL-COENZYME A  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NATIONAL RENEWABLE ENERGY LABORATORY  
STREET: 1617 Cole Blvd.  
CITY: Golden  
STATE: CO  
COUNTRY: USA  
ZIP: 80401-3393  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/418,893D

```
; FILING DATE: April 7, 1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/104,938
; FILING DATE: September 14, 1993
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: O'CONNOR, EDNA
; REGISTRATION NUMBER: 29,252
; REFERENCE/DOCKET NUMBER: MRI/NREL IR# 92-48CON
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303-231-1000
; TELEFAX: 303-231-1098
; TELEX:
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-418-893D-14

Query Match      4.0%; Score 14; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.021;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 CAYGNAAYGCGNGT 35
Db 14 CAYGNAAYGCGNGT 1

RESULT 8
US-09-167-513-8
; Sequence 8, Application US/09167513
; Patent No. 6388064
; GENERAL INFORMATION:
; APPLICANT: Conklin, Dartell C.
; APPLICANT: Blumberg, Hal
; TITLE OF INVENTION: A HUMAN 2-19 PROTEIN HOMOLOGUE, Z219A
; FILE REFERENCE: 97-63
; CURRENT APPLICATION NUMBER: US/09/167,513
; CURRENT FILING DATE: 1998-10-06
; EARLIER APPLICATION NUMBER: US 60/061,712
; EARLIER FILING DATE: 1997-10-06
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 705
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: variation
; LOCATION: (1)...(705)
; OTHER INFORMATION: z219a Degenerate polynucleotide sequence
; OTHER INFORMATION: N is any nucleotide
; US-09-167-513-8

Query Match      4.0%; Score 14; DB 4; Length 705;
Best Local Similarity 100.0%; Pred. No. 0.0064;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 173 AYCNAARAAYAY 186
Db 632 AYCNAARAAYAY 645

RESULT 9
US-08-859-183-3
; Sequence 3, Application US/08859183
; Patent No. 5827731
```

```
; GENERAL INFORMATION:
; APPLICANT: Friedrich, Thomas
; APPLICANT: Bialojan, Siegfried
; APPLICANT: Bollschweiler, Claus
; APPLICANT: Kuenast, Christoph
; TITLE OF INVENTION: No. 5827731el thrombin-inhibitory protein
; TITLE OF INVENTION: from ticks
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Keil & Weinkeuf
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.2 Mb storage
; COMPUTER: IBM AT-compatible, 80486 processor
; OPERATING SYSTEM: MS-DOS version 6.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/859,183
; FILING DATE: 20-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/415,999
; FILING DATE: 04-APR-1995
; APPLICATION NUMBER: PCT/EP92/02179
; FILING DATE: 21-SEP-1992
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-859-183-3

Query Match      3.7%; Score 13; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 AYACNGCGAYTG 197
Db 5 AYACNGCGAYTG 17

RESULT 10
US-08-447-500-13
; Sequence 13, Application US/08447500
; Patent No. 5627064
; GENERAL INFORMATION:
; APPLICANT: Hoeckstra, Merl F.
; TITLE OF INVENTION: PROTEIN KINASES
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Spensley Horn Jubas & Lubitz
; STREET: 1880 Century Park East, Suite 500
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90067
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/447,500
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/008,001
; FILING DATE:
```



```
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Wetherell Ph.D., John R.
; REGISTRATION NUMBER: 31,678
; REFERENCE/DOCKET NUMBER: PD-2458
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 455-5100
; TELEFAX: (619) 455-5110
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 23 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; IMMEDIATE SOURCE:
; CLONE: Protein Kinase
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..23
; OTHER INFORMATION: /note= "Bases designated N at
; positions 3, 6, 9, 12 and 18 are Inosine."
;
US-08-447-500-13

Query Match 3.7%; Score 13; DB 1; Length 23;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GGWSNGGWSNT 133
Db 1 GGWSNGGWSNT 13

RESULT 11
US-08-454-097-13
; Sequence 13, Application US/08454097
; Patent No. 5686412
; GENERAL INFORMATION:
; APPLICANT: Hoekstra, Merl F.
; TITLE OF INVENTION: Protein Kinases
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/454,097
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/185,359
; FILING DATE: 21-JAN-1994
; APPLICATION NUMBER: US 08/008,001
; FILING DATE: 21-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/728,783
; FILING DATE: 03-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5686412and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/31853
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 13:
```

```
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 23 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; IMMEDIATE SOURCE:
; CLONE: Protein Kinase
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..23
; OTHER INFORMATION: /note= "Bases designated N at
; positions 3, 6, 9, 12 and 18 are Inosine."
;
US-08-454-097-13

Query Match 3.7%; Score 13; DB 1; Length 23;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GGWSNGGWSNT 133
Db 1 GGWSNGGWSNT 13

RESULT 12
US-08-453-866-13
; Sequence 13, Application US/08453866
; Patent No. 5756289
; GENERAL INFORMATION:
; APPLICANT: Hoekstra, Merl F.
; TITLE OF INVENTION: PROTEIN KINASES
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Spensley Horn Jubas & Lubitz
; STREET: 1880 Century Park East, Suite 500
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90067
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,866
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/008,001
; FILING DATE: 20-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Wetherell Ph.D., John R.
; REGISTRATION NUMBER: 31,678
; REFERENCE/DOCKET NUMBER: PD-2458
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 455-5100
; TELEFAX: (619) 455-5110
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 23 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; IMMEDIATE SOURCE:
; CLONE: Protein Kinase
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..23
; OTHER INFORMATION: /note= "Bases designated N at
; positions 3, 6, 9, 12 and 18 are Inosine."
;
US-08-453-866-13
```

Query Match 3.7%; Score 13; DB 1; Length 23;  
Best Local Similarity 100.0%; Pred. No. 0.13;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 121 GGNWSNGNWSNT 133  
|||||  
Db 1 GGNWSNGNWSNT 13

## RESULT 13

US-08-185-359-13  
; Sequence 13, Application US/08185359  
; Patent No. 6060296  
; GENERAL INFORMATION:  
; APPLICANT: Hoekstra, Merl F.  
; TITLE OF INVENTION: Protein Kinases  
; NUMBER OF SEQUENCES: 57  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sears Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/185,359  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/008,001  
; FILING DATE: 21-JAN-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/728,783  
; FILING DATE: 03-JUL-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: No. 6060296and, Greta E.  
; REGISTRATION NUMBER: 35,302  
; REFERENCE/DOCKET NUMBER: 27866/31853  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 23 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; IMMEDIATE SOURCE:  
; CLONE: Protein Kinase  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..23  
; OTHER INFORMATION: /note= "Bases designated N at  
; OTHER INFORMATION: positions 3, 6, 9, 12 and 18 are Inosine."  
US-08-185-359-13

Query Match 3.7%; Score 13; DB 3; Length 23;  
Best Local Similarity 100.0%; Pred. No. 0.13;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 121 GGNWSNGNWSNT 133  
|||||  
Db 1 GGNWSNGNWSNT 13

## RESULT 14

US-08-457-192-2  
; Sequence 2, Application US/08457192  
; Patent No. 5811514  
; GENERAL INFORMATION:  
; APPLICANT: Bard, Frederique  
; APPLICANT: vednock, Theodore A.  
; APPLICANT: Keim, Pamela S.  
; TITLE OF INVENTION: NOVEL INTEGRIN BLOCKING MOLECULE  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend Hourie and Crew  
; STREET: 379 Lytton Avenue  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: US  
; ZIP: 94301  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/457,192  
; FILING DATE: HEREWITH  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/104,210  
; FILING DATE: 09-AUG-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Smith, William M.  
; REGISTRATION NUMBER: 30,223  
; REFERENCE/DOCKET NUMBER: 15270-001330  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 326-2400  
; TELEFAX: (415) 326-2422  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 72 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-457-192-2

Query Match 3.7%; Score 13; DB 1; Length 72;  
Best Local Similarity 100.0%; Pred. No. 0.089;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 148 WSNAYCCNWSNA 160  
|||||  
Db 37 WSNAYCCNWSNA 49

## RESULT 15

US-09-124-541-10  
; Sequence 10, Application US/09124541A  
; Patent No. 6229666  
; GENERAL INFORMATION:  
; APPLICANT: Morris Ph.D., Roy O.  
; TITLE OF INVENTION: A CYTOKININ OXIDASE  
; FILE REFERENCE: UMO1490  
; CURRENT APPLICATION NUMBER: US/09/124,541A  
; CURRENT FILING DATE: 1998-07-29  
; EARLIER APPLICATION NUMBER: 60/054,268  
; EARLIER FILING DATE: 1997-07-30  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 10  
; LENGTH: 1602  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:

```
NAME/KEY: variation
LOCATION: (6)
OTHER INFORMATION: a,g,c or t
FEATURE:
NAME/KEY: variation
LOCATION: (9)
OTHER INFORMATION: a,g,c or t
FEATURE:
NAME/KEY: variation
LOCATION: (12)
OTHER INFORMATION: a,g,c or t
FEATURE:
NAME/KEY: variation
LOCATION: (21)
OTHER INFORMATION: a,g,c or t
FEATURE:
NAME/KEY: variation
LOCATION: (24)
OTHER INFORMATION: a,g,c or t
FEATURE:
NAME/KEY: variation
LOCATION: (27)
OTHER INFORMATION: a,g,c or t
FEATURE:
NAME/KEY: variation
LOCATION: (30)
OTHER INFORMATION: a,g,c or t
FEATURE:
NAME/KEY: variation
LOCATION: (33)
OTHER INFORMATION: a,g,c or t
FEATURE:
NAME/KEY: variation
LOCATION: (36)
OTHER INFORMATION: a,g,c or t
FEATURE:
NAME/KEY: variation
LOCATION: (42)
OTHER INFORMATION: a,g,c or t
FEATURE:
NAME/KEY: variation
LOCATION: (48)
OTHER INFORMATION: a,g,c or t
FEATURE:
NAME/KEY: variation
LOCATION: (54)
OTHER INFORMATION: a,g,c or t
FEATURE:
NAME/KEY: variation
LOCATION: (57)
OTHER INFORMATION: a,g,c or t
FEATURE:
NAME/KEY: variation
LOCATION: (60)
OTHER INFORMATION: a,c,g or t
FEATURE:
NAME/KEY: variation
LOCATION: (63)
OTHER INFORMATION: a,g,c or t
FEATURE:
NAME/KEY: variation
LOCATION: (66)
OTHER INFORMATION: a,g,c or t
FEATURE:
NAME/KEY: variation
LOCATION: (69)
OTHER INFORMATION: a,g,c or t
FEATURE:
NAME/KEY: variation
LOCATION: (72)
OTHER INFORMATION: a,g,c or t
FEATURE:
NAME/KEY: variation
```

```
LOCATION: (75)
OTHER INFORMATION: a,g,c or t
FEATURE:
NAME/KEY: variation
LOCATION: (78)
OTHER INFORMATION: a,g,c or t
FEATURE:
NAME/KEY: variation
LOCATION: (81)
OTHER INFORMATION: a,g,c or t
FEATURE:
NAME/KEY: variation
LOCATION: (90)
OTHER INFORMATION: a,g,c or t
FEATURE:
NAME/KEY: variation
LOCATION: (93)
OTHER INFORMATION: a,g,c or t
FEATURE:
NAME/KEY: variation
LOCATION: (96)
OTHER INFORMATION: a,g,c or t
FEATURE:
NAME/KEY: variation
LOCATION: (99)
OTHER INFORMATION: a,g,c or t
FEATURE:
NAME/KEY: variation
LOCATION: (105)
OTHER INFORMATION: a,g,c or t
FEATURE:
NAME/KEY: variation
LOCATION: (108)
OTHER INFORMATION: a,g,c or t
FEATURE:
NAME/KEY: variation
LOCATION: (111)
OTHER INFORMATION: a,g,c or t
FEATURE:
NAME/KEY: variation
LOCATION: (114)
OTHER INFORMATION: a,g,c or t
FEATURE:
NAME/KEY: variation
LOCATION: (117)
OTHER INFORMATION: a,g,c or t
FEATURE:
NAME/KEY: variation
LOCATION: (120)
OTHER INFORMATION: a,g,c or t
FEATURE:
NAME/KEY: variation
LOCATION: (123)
OTHER INFORMATION: a,g,c or t
FEATURE:
NAME/KEY: variation
LOCATION: (126)
OTHER INFORMATION: a,g,c or t
FEATURE:
NAME/KEY: variation
LOCATION: (129)
OTHER INFORMATION: a,g,c or t
FEATURE:
NAME/KEY: variation
LOCATION: (135)
OTHER INFORMATION: a,g,c or t
FEATURE:
NAME/KEY: variation
LOCATION: (141)
OTHER INFORMATION: a,g,c or t
FEATURE:
NAME/KEY: variation
LOCATION: (144)
```

```
; OTHER INFORMATION: a,g,c or t
; FEATURE:
; NAME/KEY: variation
; LOCATION: (147)
; OTHER INFORMATION: a,g,c or t
; FEATURE:
; NAME/KEY: variation
; LOCATION: (153)
; OTHER INFORMATION: a,g,c or t
; FEATURE:
; NAME/KEY: variation
; LOCATION: (159)
; OTHER INFORMATION: a,c,g or t
; FEATURE:
; NAME/KEY: variation
; LOCATION: (162)
; OTHER INFORMATION: a,g,c or t
; FEATURE:
; NAME/KEY: variation
; LOCATION: (165)
; OTHER INFORMATION: a,g,c or t
; FEATURE:
; NAME/KEY: variation
; LOCATION: (168)
; OTHER INFORMATION: a,g,c or t
; FEATURE:
; NAME/KEY: variation
; LOCATION: (171)
; OTHER INFORMATION: a,g,c or t
; FEATURE:
; NAME/KEY: variation
; LOCATION: (174)
; OTHER INFORMATION: a,g,c or t
; FEATURE:
; NAME/KEY: variation
; LOCATION: (177)
; OTHER INFORMATION: a,g,c or t
```

Query Match 3.7%; Score 13; DB 4; Length 1602;

Best Local Similarity 100.0%; Pred. No. 0.032;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 96 NGCNWSNGTNTN 108

|||||

Db 1050 NGCNWSNGTNTN 1062

Search completed: November 7, 2002, 04:59:56  
Job time : 72 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 23:36:55 ; Search time 2767 Seconds  
(without alignments)  
3691.755 Million cell updates/sec

Title: US09897438BK-2  
Perfect score: 351  
Sequence: 1 GARGARTGYGNCACATAT.....AYGARGCNCYTGCGGCNTN 351-

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_hgt.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
1	249.4	71.1	11673	6	AX305377	AX305377 Sequence
2	249.4	71.1	11673	10	MMU24703	U24703 Mus musculus
3	240.6	68.5	11187	10	AB049473	AB049473 Rattus no
4	233.8	66.6	11580	6	AX410790	AX410790 Sequence
5	233.8	66.6	11580	9	HSU79716	U79716 Human reeli
6	233.8	66.6	11580	11	G30936	G30936 SWS52926 Er
7	233.8	66.6	11580	11	G30938	G30938 SWS53176 Er
8	172.8	49.2	10634	5	AF090441	AF090441 Gallus ga
c 9	96.6	27.5	163985	10	AC121878	AC121878 Mus muscu
10	96.6	27.5	183156	2	AC023062	AC023062 Mus muscu
11	92.6	26.4	126130	2	AC095877	AC095877 Rattus no
c 12	92.6	26.4	202764	2	AC128022	AC128022 Rattus no
c 13	91.6	26.1	183641	2	AC124933	AC124933 Rattus no
c 14	90.2	25.7	93163	9	HSAC000121	AC000121 Human BAC
c 15	53.4	15.2	185996	2	AC041023	AC041023 Homo sapi
16	49.6	14.1	158	4	AF232904	AF232904 Bos tauru
17	48	13.7	125020	9	AF429315	AF429315 Homo sapi
c 18	47	13.4	125020	9	AF429315	AF429315 Homo sapi
19	45.2	12.9	146383	2	AC116367	AC116367 Oryza sat
20	43.8	12.5	161034	9	AC015771	AC015771 Homo sapi
21	43.8	12.5	161707	2	AC015772	AC015772 Homo sapi
c 22	43.6	12.4	115666	2	AC105744	AC105744 Oryza sat
c 23	42.6	12.1	115666	2	AC105744	AC105744 Oryza sat
c 24	42	12.0	68341	2	AC105584	AC105584 Rattus no
c 25	41.8	11.9	173926	2	MM454K24	AL160232 Mus muscu
c 26	41.6	11.9	104695	9	HSUJ96108	AL121879 Human DNA
c 27	41.6	11.9	162355	2	AC092957	AC092957 Homo sapi
c 28	41.6	11.9	176186	9	AC006925	AC006925 Homo sapi
c 29	41.6	11.9	188688	2	AL833801	AL833801 Mus muscu
c 30	41.2	11.7	153645	9	AC025426	AC025426 Homo sapi
31	41.2	11.7	165110	9	AL450311	AC025426 Homo sapi
c 32	40.6	11.6	33556	2	AC020474	AL450311 Human DNA
33	40.6	11.6	172105	3	AC104625	AC020474 Drosophil
c 34	40.6	11.6	183334	2	AC106174	AC104625 Drosophil
c 35	40.6	11.6	302915	3	AE003489	AC106174 Rattus no
c 36	40.2	11.5	1141	6	AX083744	AE003489 Drosophil
37	40.2	11.5	1246	6	AX164174	AX083744 Sequence
c 38	39.8	11.3	146383	2	AC116367	AX164174 Sequence
c 39	39.6	11.3	173341	2	AC021954	AC116367 Oryza sat
c 40	39.4	11.2	157197	2	AC099042	AC021954 Homo sapi
c 41	39.4	11.2	192195	2	AC099620	AC099042 Oryza sat
c 42	39.2	11.2	3461	3	DROM1RADH	AC099620 Mus muscu
c 43	39.2	11.2	7218	6	I66494	M60998 Drosophila
c 44	39.2	11.2	156206	2	AC105500	I66494 Sequence 14
c 45	39.2	11.2	293184	2	AC079487	AC105500 Rattus no
						AC079487 Mus muscu

ALIGNMENTS

RESULT 1  
AX305377  
LOCUS AX305377 11673 bp DNA linear PAT 11-DEC-2001  
DEFINITION Sequence 128 from Patent WO0188188.  
ACCESSION AX305377  
VERSION AX305377.1 GI:17644926  
KEYWORDS house mouse.  
SOURCE Mus musculus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
REFERENCE 1  
AUTHORS Ishikawa,K., Asai,S., Takahashi,Y., Nagata,T. and Ishii,Y.  
TITLE Method for examining ischemic conditions  
JOURNAL Patent: WO 0188188-A 128 22-NOV-2001;









VVSTDAVAVNEDSLQIDFAASCSVTDSYAEILEYSVDLGLSWHLVRDCLPTNVES  
RYHLORILYDFNKVCRITLPLPPYRTOATFRWHPQAPDKQOTWADNVIYIGDG  
CIDMCSHGRCIOGNCVCEOMGLYCDOPETSLPTOLKDNFNAESPSONNLVNGGK  
LSTVCASAGMALHFGSGCSRLLYTVDLNLNAEIQFYFMTCGLITPNRQGVLL  
EYSVNGITWNLMEIFYDYISKAPGVNILLPPDAKEIATFRWQPRHDLGDLONDA  
IDNVLSSGADQETVMDLTFSSAPQHERSPADAGPVGRIAFDMFMDKTSVNEHL  
FHDDCTVERFCDSPDGMVLCVSHGSDREVAVTHDLTPTGWIQFKISVGCKVSEKIA  
QNDIHWQYSTDFGVSNVLPQCLPADPKCSGVSQSPVFFPTKWKRTIYPLPESLV  
GNPVREFYOKYSDMOWAIDNFIYVGLPGCLDNCRGHGDCLREOCICDPYSGPNCYLTH  
TLKFLKERDSEIEIKPDLWMSLEGGSTCTCEGLAEDIALYFGGTVRQAVQTDLDL  
RGAKFYQYRGSENNMTSCHRPICKRKEVLDDYSTDGGITWTLHEMDYQIYSVR  
HDXLPELALNTLRWQPFVINGVSVVERAOWALDNLIGAEINPSQLVD  
TEYDEPTHEENMSFYNAVRTAGFCGNPSFLHWPNNKKDKTHNALSSRELIIQCY  
MMQFKLVGCEATSCDGLHSMVLEYTKDARSDSQWLVQTCPLSPSSNSIGCSPFOFHE  
ATYNSVNSKKRITIQLPDHSVSSATOPRWIOKGEETEKOSWADHWYIGACPKL  
CSGGCTTGATCICDESEFGDDCSVESDLPYIKDNFESARVTANWETIOGGVIG  
SCGGLAPTAHGSDFVNGCQIQAATKPLDLFRASKIMFYLOIGMSQSDSCNSDL  
GPHAVDAVLLQYVSVNNGTHTWVIAHQPKDFQAKQVSYNVPLEARMKGVLLRWQ  
RINTGHDQWALDHVEVLVSTYTRKQYMMNFRSQHGLRHFYNNRRSLRRYP  
BASE COUNT 3014 a 2696 c 2753 g 3116 t 1 others  
ORIGIN

Query Match 66.6%; Score 233.8; DB 9; Length 11580;  
Best Local Similarity 56.0%; Pred. No. 4.5e-52;  
Matches 196; Conservative 74; Mismatches 80; Indels 0; Gaps 0;  
QY 1 GARCARTGYGNACNATNATCCAYGNAAYCNCNTNACNTTYTGYGARCCTNATYGGNCCN 60  
DB 860 GAACAGTGTGGCGGATATGTCATGCAATGCCGTACCCTCTGTGAACCATATGGCCCA 919  
QY 61 MNGARYTNACNACNACNTGYTNAAYACNACNACNACNACNACNTNATYTCARTYWSNATN 120  
DB 920 CGAAGTGTATCCACAGGCGCTTATACACACACACCTCTGTCTCCATTTCCATT 979  
QY 121 GGNWSNGNSNTGYMNTYWSNTAYWSNGAYCCNWSNATNACNACNACNACNATYGCNNAAR 180  
DB 980 GGGTCAGGTTTCATGCTGCTTATTCAGACCCAGCATCATCTGTATTATGCGCAAG 1039  
QY 181 AAYAAACNACNACNACNACNACNACNACNACNACNACNACNACNACNACNACNACNACN 240  
DB 1040 ANTAACCTCTGGGACTGGATTCAGTAGAGAAATAGAGCCCTTCCATGTGAGCACA 1099  
QY 241 GTNATNACNATNTAYTNCNGARGCNAARGGNGARWSNACNATYTCARTYTCARTGG 300  
DB 1100 ATCATCCATATCTCTACCTTCTGAGGAGCCCAAGGAGAGATGTCATTTTCAGTGG 1159  
QY 301 AARCAAGYNSNTNMGNGTNGNGARGTAYGARGCNGTYTGGGCGNYT 350  
DB 1160 AAGCAGGAAATCTCTGCTAGGTGAGTGTATGATGAGCCCTGCTGGGCGTT 1209

RESULT 6  
G30936  
LOCUS  
DEFINITION 11580 bp DNA linear STS 28-SEP-1998  
G30936 swSS2926 Eric D. Green Homo sapiens STS genomic, sequence tagged  
site.  
ACCESSION G30936  
VERSION G30936.1 GI:1923209  
KEYWORDS STS.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 11580)  
Bouffard,G.G., Iyer,L.M., Idol,J.R., Braden,V.V., Cunningham,A.F.,  
Weintraub,L.A., Mohr-Tidwell,R.M., Peluso,D.C., Fulton,R.S.,  
Leckie,M.P. and Green,E.D.  
A collection of 1814 human chromosome 7-specific STSs  
Genome Res. 7 (1), 59-64 (1997)  
97189344  
MEDLINE 9037602  
PUBMED  
REFERENCE 2 (bases 1 to 11580)  
AUTHORS Green,E.D.

TITLE Human chromosome 7 STSs (1997)  
JOURNAL Unpublished (1997)  
COMMENT On Apr 3, 1997 this sequence version replaced gi:1706935.  
Synonyms: RELN  
GDB\_DSEG: RELN  
Contact: Eric D. Green  
Genome Technology Branch  
National Human Genome Research Institute/NIH  
49 Convent Dr., MSC4431, Bldg. 49, Rm. 2A08, Bethesda, MD 20892  
Tel: 3014020201  
Fax: 3014024735  
Email: egreen@hghri.nih.gov  
Primer A: CTCACCTTCCTGAGGACGCCAA  
Primer B: AAGCCAGTTGCCGTCTCCACTG  
STS size: 183  
PCR profile:  
Presoak: 0 degrees C for 0.00 minute(s)  
Denaturation: 92 degrees C for 0.17 minute(s)  
Annealing: 55 degrees C for 1.00 minute(s)  
Polymerization: 72 degrees C for 1.00 minute(s)  
PCR Cycles: 35  
Thermal Cycler: PerkinElmer 9600  
Protocol:  
Template: 30-100 ng  
Primer: each 1 uM  
dNTPs: each 200 uM  
Taq Polymerase: 0.05 units/ul  
Total Vol: 10 ul  
Buffer:  
MgCl2: 1.5 mM  
KCl: 100 mM  
Tris-HCl: 10 mM  
NH4Cl: 5 mM  
pH: 8.6

This STS was developed from sequence determined by another investigator. See GenBank record: U79716 For additional information about the NHGRI chromosome 7 mapping project, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>. Also see Genomics 11:548-64 (1991) [NUID=92128937].

FEATURES  
source 1..11580  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="7"  
gene 1..11580  
/gene="RELN"  
STS 1112..1294  
/gene="RELN"  
primer\_bind 1112..1134  
/gene="RELN"  
primer\_bind complement(1272..1294)  
BASE COUNT 3014 a 2696 c 2753 g 3116 t 1 others  
ORIGIN

Query Match 66.6%; Score 233.8; DB 11; Length 11580;  
Best Local Similarity 56.0%; Pred. No. 4.5e-52;  
Matches 196; Conservative 74; Mismatches 80; Indels 0; Gaps 0;  
QY 1 GARCARTGYGNACNATNATCCAYGNAAYCNCNTNACNTTYTGYGARCCTNATYGGNCCN 60  
DB 860 GAACAGTGTGGCGGATATGTCATGCAATGCCGTACCCTCTGTGAACCATATGGCCCA 919  
QY 61 MNGARYTNACNACNACNTGYTNAAYACNACNACNACNACNACNACNATYTCARTYWSNATN 120  
DB 920 CGAAGTGTATCCACAGGCGCTTATACACACACACCTCTGTCTCCATTTCCATT 979  
QY 121 GGNWSNGNSNTGYMNTYWSNTAYWSNGAYCCNWSNATNACNACNACNACNATYGCNNAAR 180  
DB 980 GGGTCAGGTTTCATGCTGCTTATTCAGACCCAGCATCATCTGTATTATGCGCAAG 1039





[illegible]

Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G.,  
 Oraqunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,  
 Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G.,  
 Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,  
 Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,  
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
 Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,  
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
 Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,  
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstein, G., and Gibbs, R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 126130)  
 Worley, K.C.  
 Direct Submission  
 Submitted (17-SEP-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 126130)  
 Worley, K.C.  
 Direct Submission  
 Submitted (11-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 10, 2002 this sequence version replaced gi:20975941.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GDRS  
 Center clone name: CH230-10G13  
 ----- Summary Statistics  
 Sequencing vector: Plasmid  
 Chemistry: Dye-terminator Big Dye: 100% of reads  
 Assembly program: Phrap: version 0.990329  
 Consensus quality: 70934 bases at least Q40  
 Consensus quality: 75985 bases at least Q30  
 Consensus quality: 80731 bases at least Q20  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 59 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

12811  
 14181  
 14280: contig of 1370 bp in length  
 16048: gap of unknown length  
 16048: contig of 1768 bp in length  
 16148: gap of unknown length  
 17569: contig of 1421 bp in length  
 17570: gap of unknown length  
 17570: contig of 1151 bp in length  
 18820: gap of unknown length  
 18920: contig of 1289 bp in length  
 20209: gap of unknown length  
 20309: contig of 1324 bp in length  
 21633: gap of unknown length  
 21733: contig of 1536 bp in length  
 23269: gap of unknown length  
 23369: contig of 1278 bp in length  
 24647: gap of unknown length  
 24747: contig of 1053 bp in length  
 25800: gap of unknown length  
 25900: contig of 1117 bp in length  
 27017: gap of unknown length  
 27117: contig of 2066 bp in length  
 29183: gap of unknown length  
 29283: contig of 1032 bp in length  
 30315: gap of unknown length  
 30415: contig of 1671 bp in length  
 32086: gap of unknown length  
 32186: contig of 1209 bp in length  
 33395: gap of unknown length  
 33495: contig of 2168 bp in length  
 35663: gap of unknown length  
 35763: contig of 2281 bp in length  
 38044: gap of unknown length  
 38144: contig of 1701 bp in length  
 39845: gap of unknown length  
 39945: contig of 2037 bp in length  
 41982: gap of unknown length  
 42082: contig of 1661 bp in length  
 43743: gap of unknown length  
 43843: contig of 2061 bp in length  
 45904: gap of unknown length  
 46004: contig of 1663 bp in length  
 47667: gap of unknown length  
 47767: contig of 2066 bp in length  
 49833: gap of unknown length  
 49933: contig of 1501 bp in length  
 51434: gap of unknown length  
 51534: contig of 2244 bp in length  
 53778: gap of unknown length  
 53878: contig of 1674 bp in length  
 55552: gap of unknown length  
 55652: contig of 2018 bp in length  
 57670: gap of unknown length  
 57770: contig of 1860 bp in length  
 59630: gap of unknown length  
 59631: contig of 1417 bp in length  
 61147: gap of unknown length  
 61247: contig of 2748 bp in length  
 63995: gap of unknown length  
 64095: contig of 2199 bp in length  
 66294: gap of unknown length  
 66394: contig of 2400 bp in length  
 68794: gap of unknown length  
 68894: contig of 2639 bp in length  
 71533: gap of unknown length  
 71633: contig of 1836 bp in length  
 73469: gap of unknown length  
 73569: contig of 3485 bp in length  
 77054: gap of unknown length  
 77154: contig of 1851 bp in length  
 79005: gap of unknown length  
 79105: contig of 2189 bp in length  
 81294: gap of unknown length  
 81394: contig of 2470 bp in length  
 83864: gap of unknown length

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

```

* 83865 83964: gap of unknown length
* 83965 86039: contig of 2075 bp in length
* 86040 86139: gap of unknown length
* 86140 88646: contig of 2507 bp in length
* 88647 88745: gap of unknown length
* 88747 91823: contig of 3077 bp in length
* 91824 91923: gap of unknown length
* 94369 94469: contig of 2446 bp in length
* 94370 96886: gap of unknown length
* 94470 96886: contig of 2417 bp in length
* 96887 96986: gap of unknown length
* 96987 100679: contig of 3693 bp in length
* 100680 100779: gap of unknown length
* 100780 104171: contig of 3392 bp in length
* 104172 104271: gap of unknown length
* 104272 106760: contig of 2489 bp in length

Query Match      26.4%; Score 92.6; DB 2; Length 126130;
Best Local Similarity 57.5%; Pred. No. 1.6e-13;
Matches 77; Conservative 29; Mismatches 28; Indels 0; Gaps 0;

QY 217 MNGCNCNWSNAAYGTNWSNACNGTATNCAYATNTNTAYTNTCCGARGAGCNAAR 276
      :|||:::||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 92827 AGACCCCTTCCACGTAAGCACCATCATCCATCTCTACCTCTCTGAGGACGCCAAA 92886*
      :|||:::||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

QY 277 GNGARWSNNTCARTTYCARTGGAARCARGAYWSNTYNTNMGNTGNGGARGTNTAYGAR 336
      :|||:::||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 92887 GGGGAGAACGTGCAGTCCAGTCCAGTGAACAGACAGCACCTCGCTGTGGGTAGTACGAA 92946
      :|||:::||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

QY 337 GCNTGTGGGCGNYT 350
      :|||:::||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 92947 GCCTGCTGGGCCCT 92960
      :|||:::||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

```

```

RESULT 12
AC128022/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-525L20, *** SEQUENCING IN PROGRESS
***, 97 unordered pieces.
AC128022
AC128022.1 GI:21908605
HTG: HTGS_PHASE1.
Rattus norvegicus.
Rattus norvegicus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 202764)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbosa,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Devilla,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Fallis,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgeson,A., Hogue,M., Holloway,C., Hollins,B.,
Homsí,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudan,S.,
Karlssoon,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratoch,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louseged,H.,
Lozadó,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,

```

```

Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Fu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I.,
Sodergren,E., Sonaika,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansley,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 202764)
REFERENCE
Worley,K.C.
Direct Submission
Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: KAIY
Center clone name: CH230-525L20
----- Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 113069 bases at least Q40
Consensus quality: 119970 bases at least Q30
Consensus quality: 125656 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 97 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1244: contig of 1244 bp in length
* 1245 1344: gap of unknown length
* 1345 2346: contig of 1002 bp in length
* 2347 2447: gap of unknown length
* 2447 3813: contig of 1367 bp in length
* 3814 3913: gap of unknown length
* 3914 5101: contig of 1188 bp in length
* 5102 5201: gap of unknown length
* 5202 6724: contig of 1523 bp in length
* 6725 6824: gap of unknown length
* 6825 8048: contig of 1224 bp in length
* 8049 8148: gap of unknown length
* 8149 9212: contig of 1064 bp in length
* 9213 9312: gap of unknown length
* 9313 10432: contig of 1120 bp in length
* 10433 10532: gap of unknown length
* 10533 11890: contig of 1358 bp in length
* 11891 11990: gap of unknown length
* 11991 13068: contig of 1078 bp in length
* 13069 13168: gap of unknown length
* 13169 14486: contig of 1318 bp in length
* 14487 14586: gap of unknown length
* 14587 15936: contig of 1410 bp in length
* 15937 16096: gap of unknown length

```



```

Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savary,G.,
Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 183641)
Worley,K.C.
Direct Submission
Submitted (20-JUN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 183641)
Worley,K.C.
Direct Submission
Submitted (31-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 24, 2002 this sequence version replaced gi:21490053.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: KAVI
Center clone name: CH230-466N17
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 111724 bases at least Q40
Consensus quality: 120970 bases at least Q30
Consensus quality: 128458 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 88 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1107: contig of 1107 bp in length
* 1108: gap of unknown length
* 1208: contig of 1253 bp in length
* 2461: 2560: gap of unknown length
* 2561: 4004: contig of 1444 bp in length
* 4005: 4104: gap of unknown length
* 4105: 5382: contig of 1278 bp in length
* 5383: 5482: gap of unknown length
* 5483: 6874: contig of 1392 bp in length
* 6875: 6975: gap of unknown length
* 6975: 8071: contig of 1097 bp in length
* 8072: 8171: gap of unknown length
* 8172: 9478: contig of 1307 bp in length
* 9479: 9578: gap of unknown length
* 9579: 11061: contig of 1483 bp in length
* 11062: 11161: gap of unknown length
*
* 11162: 12742: contig of 1581 bp in length
* 12743: gap of unknown length
* 12843: contig of 1662 bp in length
* 14504: gap of unknown length
* 14604: gap of unknown length
* 15834: contig of 1130 bp in length
* 15835: gap of unknown length
* 16958: contig of 1124 bp in length
* 16959: gap of unknown length
* 17059: contig of 1646 bp in length
* 18804: gap of unknown length
* 18805: contig of 1616 bp in length
* 20520: gap of unknown length
* 20521: contig of 1128 bp in length
* 21649: gap of unknown length
* 21749: contig of 1429 bp in length
* 23177: gap of unknown length
* 23278: contig of 1483 bp in length
* 24761: gap of unknown length
* 24861: contig of 1899 bp in length
* 26760: gap of unknown length
* 26860: contig of 1440 bp in length
* 28300: gap of unknown length
* 28400: contig of 1193 bp in length
* 29593: gap of unknown length
* 29693: contig of 1178 bp in length
* 30871: gap of unknown length
* 30971: contig of 1672 bp in length
* 32643: gap of unknown length
* 32743: contig of 1644 bp in length
* 34387: gap of unknown length
* 34487: contig of 1578 bp in length
* 36065: gap of unknown length
* 36165: contig of 1234 bp in length
* 37399: gap of unknown length
* 37499: contig of 2110 bp in length
* 39609: gap of unknown length
* 39709: contig of 1772 bp in length
* 41481: gap of unknown length
* 41581: contig of 1179 bp in length
* 42760: gap of unknown length
* 42860: contig of 1852 bp in length
* 44712: gap of unknown length
* 44812: contig of 1397 bp in length
* 46209: gap of unknown length
* 46309: contig of 1459 bp in length
* 47268: gap of unknown length
* 47668: contig of 1360 bp in length
* 49228: gap of unknown length
* 49328: contig of 1470 bp in length
* 49328: gap of unknown length
* 50798: contig of 1031 bp in length
* 50898: contig of 1031 bp in length
* 51929: gap of unknown length
* 52029: contig of 1077 bp in length
* 53106: gap of unknown length
* 53206: contig of 1341 bp in length
* 54547: gap of unknown length
* 54647: contig of 1174 bp in length
* 55821: gap of unknown length
* 55921: contig of 1715 bp in length
* 57636: gap of unknown length
* 57736: contig of 1207 bp in length
* 58943: gap of unknown length
* 59043: contig of 1299 bp in length
* 60342: gap of unknown length
* 60442: contig of 1252 bp in length
* 61694: gap of unknown length
* 61794: contig of 1522 bp in length
* 63316: gap of unknown length
* 63416: contig of 1156 bp in length
* 64572: gap of unknown length
* 64672: contig of 1412 bp in length
* 66084: gap of unknown length
* 66184: contig of 1333 bp in length

```





repeat_region	/rpt_family="L1" 24613. .24654	/note="match to Human cDNA, 5' end H56746 (NID:1005390)"
repeat_region	/rpt_family="L1" 25141. .25180	
repeat_region	/rpt_family="L1" 26470. .26761	
misc_feature	/rpt_family="ALU" complement(26762. .26936)	
repeat_region	/gene="WUGSC:H_RG249A12.1" /note="match to Human cDNA, 5' end H56746 (NID:1005390)"	
exon	26937. .26989 /rpt_family="L1" 27107. .27238	/note="GRAIL prediction, score = 82"
repeat_region	/evidence=not_experimental 27993. .28465	
repeat_region	/rpt_family="L1" 28476. .28496	
repeat_region	/rpt_family="L1" complement(29663. .29903)	
exon	/rpt_family="ALU" complement(32793. .33032)	
repeat_region	/gene="WUGSC:H_RG249A12.1" /note="GRAIL prediction, score = 90"	
repeat_region	/evidence=not_experimental complement(33462. .33752)	
repeat_region	/rpt_family="ALU" 37975. .37994	
repeat_region	/rpt_family="L1" 40278. .40311	
repeat_region	/rpt_family="L1" complement(41921. .42075)	
repeat_region	/rpt_family="ALU" complement(42447. .42734)	
exon	/rpt_family="ALU" 43126. .43185	/note="GRAIL prediction, score = 95"
repeat_region	/evidence=not_experimental complement(43411. .43680)	
repeat_region	/rpt_family="ALU" 45704. .45991	
misc_feature	/rpt_family="ALU" 46064. .46434	/note="match to Human cDNA H60097, 5' end (NID:g1012929)"
repeat_region	complement(46604. .46897)	
misc_feature	/rpt_family="ALU" complement(47023. .47412)	
repeat_region	/gene="WUGSC:H_RG249A12.1" /note="match to Human cDNA, 3' end H59444 (NID:g1012276)"	
repeat_region	58125. .58163 /rpt_family="L1"	
repeat_region	58604. .58758 /rpt_family="L1"	
repeat_region	59743. .59765 /rpt_family="L1"	
repeat_region	60210. .60516 /rpt_family="ALU"	
exon	complement(63051. .63146)	
repeat_region	/gene="WUGSC:H_RG249A12.1" /note="GRAIL prediction, score = 100"	
repeat_region	/evidence=not_experimental 64242. .64284	
repeat_region	/rpt_family="L1" 64445. .64496	
repeat_region	/rpt_family="L1" complement(67753. .68011)	
repeat_region	/rpt_family="L1" 67986. .68889	
exon	/rpt_family="L1" complement(70725. .70841)	
	/gene="WUGSC:H_RG249A12.1" /note="GRAIL prediction, score = 82"	
	/evidence=not_experimental	

[illegible]



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 03:28:26 ; Search time 2214 Seconds  
(without alignments)  
2567.577 Million cell updates/sec

Title: US09897438BK-2

Perfect score: 351

Sequence: 1 GARCARTGYGNCATNAT.....AYGARGCNTGYTGGCNYTN 351

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_othr:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	12	3.4	321	17 A0226184	A0226184 HS.2022_B
2	12	3.4	377	14 C97776	C97776 C97776 Rice
3	12	3.4	549	17 A0674732	A0674732 HS.5485_B
c 4	12	3.4	831	12 BE705775	BE705775 Sc02_04b1
5	12	3.4	1714	17 AG029608	AG029608 Pan trogl
c 6	11	3.1	202	14 W20272	W20272 zb43h08.r1

7	11	3.1	226	14	N65130
8	11	3.1	333	14	C74396
c 9	11	3.1	361	13	BJ073290
10	11	3.1	381	14	F23596
c 11	11	3.1	408	9	AL045493
12	11	3.1	437	9	AI997638
13	11	3.1	530	9	AI562717
c 14	11	3.1	604	17	FR0043054
15	11	3.1	613	13	BJ503599
16	11	3.1	634	13	BI959766
17	11	3.1	1051	17	CNS008E4
c 18	11	3.1	1068	17	CNS05D8A
c 19	11	3.1	1071	13	BG985678
20	11	3.1	1533	17	AG107622
c 21	10	2.8	145	10	AV849260
c 22	10	2.8	148	14	T49675
c 23	10	2.8	190	13	BJ467155
c 24	10	2.8	193	14	T63392
c 25	10	2.8	222	17	B04666
c 26	10	2.8	253	9	AU210723
c 27	10	2.8	254	14	T02496
c 28	10	2.8	257	10	BB157379
c 29	10	2.8	267	13	BJ438963
30	10	2.8	278	13	BJ368840
c 31	10	2.8	283	13	BJ377516
32	10	2.8	300	9	AV199737
33	10	2.8	300	13	BJ135481
c 34	10	2.8	315	14	F08819
c 35	10	2.8	316	14	W03213
c 36	10	2.8	316	14	W03213
37	10	2.8	317	13	BJ366180
c 38	10	2.8	322	13	BJ399800
c 39	10	2.8	323	9	AA681046
c 40	10	2.8	324	14	D32970
c 41	10	2.8	378	14	T34001
c 42	10	2.8	387	12	BG520970
c 43	10	2.8	389	17	FR0028242
c 44	10	2.8	393	13	BJ131990
c 45	10	2.8	400	9	AL360472

#### ALIGNMENTS

RESULT 1  
A0226184/c  
LOCUS  
DEFINITION  
HS\_2022\_B1\_E12\_T7 CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2022 Col=23 Row=J, DNA sequence.  
ACCESSION  
A0226184  
VERSION  
A0226184.1  
KEYWORDS  
GSS.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 321)  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.  
Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
JOURNAL  
MEDLINE  
COMMENT  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallaceu.washington.edu  
Sequence Tagged Connector  
Plate: 2022 Row: J Column: 23

```

Class: BAC ends
High quality sequence stop: 321.
Location/Qualifiers
1..321
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2022 Col=23 Row=J"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 93 a 63 c 56 g 102 t 7 others
ORIGIN

Query Match 3.4%; Score 12; DB 17; Length 321;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 NACNATNGCA 23
|||||
Db 247 NACNATNGCA 236

RESULT 2
C97776
LOCUS C97776 377 bp mRNA linear EST 04-APR-2002
DEFINITION C97776 Rice callus Oryza sativa (japonica cultivar-group) cDNA
ACCESSION clone C62743_12, mRNA sequence.
VERSION C97776.1 GI:3760522
KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group).
ORGANISM Oryza sativa (japonica cultivar-group).
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
1 (bases 1 to 377)
Sasaki, T. and Minobe, Y.
Rice cDNA from callus
Unpublished (1994)
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasaki@abrr.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/
PROJECT = "RGP"
POLYA=No.
Location/Qualifiers
1..377
/organism="Oryza sativa (japonica cultivar-group)"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="C62743_12"
/clone_lib="Rice callus"
/note="Vector: pBluescript II SK+; Site_1: SalI; Site_2:
NotI; cDNA prepared from rice callus mRNAs by using
oligo(dT) as a primer and ligating to the SalI-NotI site
of pBluescript II SK+ phagemid."
BASE COUNT 95 a 76 c 83 g 110 t 13 others
ORIGIN

Query Match 3.4%; Score 12; DB 14; Length 377;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 ACNACNACNGCN 99
|||||
Db 173 ACNACNACNGCN 184

RESULT 3

```

```

A0674732
LOCUS HS_5485_B1_G05_T7A RPCI-11 Human Male BAC Library Homo sapiens
DEFINITION genomic clone Plate=1061 Col=9 Row=N, DNA sequence.
ACCESSION A0674732
VERSION A0674732.1 GI:5207478
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 549)
Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
Hood, L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Resear h Genetics (Info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 1061 row: N column: 9
Seq primer: T7
Class: BAC ends
High quality sequence stop: 549.
Location/Qualifiers
1..549
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=1061 Col=9 Row=N"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"
BASE COUNT 158 a 112 c 111 g 159 t 9 others
ORIGIN

Query Match 3.4%; Score 12; DB 17; Length 549;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 NACNACNACNTG 80
|||||
Db 28 NACNACNACNTG 39

RESULT 4
BE705775/c
LOCUS BE705775 831 bp mRNA linear EST 12-SEP-2000
DEFINITION Sc02_04b10_A Sc02_AAFc_EC0RC_cold_stressed_winter_rye_seedlings
ACCESSION BE705775
VERSION BE705775.1 GI:10094040
KEYWORDS EST.
SOURCE rye.
ORGANISM Secale cereale
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
; Triticeae; Secale.

```

```

REFERENCE
AUTHORS Singh,J.A., Piche,C., Couroux,P., De Moors,A., Harris,L.J., Hattori
        J.I., Ouellet,T., Robert,L.S., Sprott,D. and Tinker,N.A.
TITLE Expressed Sequence Tags from Cold-Stressed Winter Rye Seedlings
JOURNAL Unpublished (2000)
COMMENT Contact: Singh,J.A.
        Eastern Cereal and Oilseed Research Centre
        Agriculture and Agri-food Canada
        KW Neatby Bldg., Central Experimental Farm, Ottawa, Ontario, K1A
        OC6, Canada
        Tel: (613) 759-1662
        Fax: (613) 759-1701
        Email: singhja@em.agr.ca.

FEATURES
source
1. .831
   Location/Qualifiers
   /organism="Secale cereale"
   /cultivar="Puma (winter rye)"
   /db_xref="taxon:4550"
   /clone="Sc02.04b10"
   /clone_lib="Sc02_AAFc_ECORC_cold_stressed_winter_rye_seedlings"
   /tissue_type="leaf, crown"
   /dev_stage="seedling three-leaf stage"
   /notes="Vector: Bluescript SK+/XhoI-EcoRI; Site.1: Eco RI;
   Site.2: Xho I; Sampled three-leaf seedlings treated for
   one week at 20C, 12 hrs light/day. Library made with
   Stratagene UNIZAP XR kit(not packaged). cDNA is directly
   ligated into SK+/XhoI-EcoRI, then electroporated into
   TOP10 cells (Invitrogen)."
```

BASE COUNT 78 a 180 c 211 g 264 t 98 others

ORIGIN

```

Query Match 3.4%; Score 12; DB 12; Length 831;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 TNACNACNACNT 79
|||||
Db 705 TNACNACNACNT 694
|||||

RESULT 5
AG029608 1714 bp DNA linear GSS 01-NOV-2001
Pan troglodytes DNA, clone: PTB-001L22.R, genomic survey sequence.
AG029608
AG029608.1 GI:16556481
GSS.
SOURCE
Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
BAC Library clone:PTB-001L22.R.

ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
BAC end sequences of Library PTB
Unpublished
2 (bases 1 to 1714)
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chmpbes@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: M13Rev
LIBRARY
```

```

Vector : PKS145
R.Site 1 : SacI
R.Site 2 : SacI
Location/Qualifiers
1. .1714
   /organism="Pan troglodytes"
   /db_xref="taxon:9598"
   /clone="PTB-001L22.R"
   /sex="male"
   /cell_type="lymphoblast"
   /clone_lib="PTB Chimpanzee Male BAC Library"
   /clone_type="PTB" 472 g 325 t 93 others

BASE COUNT 527 a 297 c 472 g 325 t 93 others

ORIGIN

Query Match 3.4%; Score 12; DB 17; Length 1714;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 TNACNACNACNT 79
|||||
Db 1616 TNACNACNACNT 1627
|||||

RESULT 6
W20272/c
LOCUS W20272 202 bp mRNA linear EST 03-MAY-1996
DEFINITION ZB43H08.rl Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone
IMAGE:306399 5' similar to WP:T09A5.6 CE01087 ; mRNA sequence.
W20272
ACCESSION W20272.1 GI:1296206
VERSION W20272.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
        M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
        Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
        R., Williamson,A., Wohlmann,P. and Wilson,R.
        The WashU-Merck EST Project
        Contact: Wilson RK
        Washington University School of Medicine
        4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
        Tel: 314 286 1800
        Fax: 314 286 1810
        Email: est@watson.wustl.edu
        This clone is available royalty-free through LLNL ; contact the
        IMAGE Consortium (info@image.llnl.gov) for further information.
        Trace considered overall poor quality
        Possible reversed clone: similarity on wrong strand
        Seq primer: mob.REGA+ET
        High quality sequence stop: 1.

FEATURES
source
1. .202
   /organism="Homo sapiens"
   /db_xref="GDB:124981"
   /db_xref="taxon:9606"
   /clone="IMAGE:306399"
   /clone_lib="Soares_fetal_lung_NbHL19W"
   /dev_stage="19 weeks"
   /lab_host="DH10B (ampicillin resistant)"
   /note="Organ: lung; Vector: pT73D (Pharmacia) with a
   modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
   strand cDNA was primed with a Not I - oligo(dT) primer
   [5'-TCTTACCAATCTGAAGTGGAGCGCCGAATTTTTTTTTTTT-3'],
   double-stranded cDNA was size selected, ligated to Eco RI
   adapters (Pharmacia), digested with Not I and cloned into
   the Not I and Eco RI sites of a modified pT73 vector
   (Pharmacia). Library went through one round of
   normalization to a Cot = 5. Library constructed by Bento
   Soares and M.Fatima Bonaldo. This library was constructed
```

68	TNACNACNACN	78
54	TNACNACNACN	64

FEATURES  
SOURCE

1. .361





Gorgone,G., Burns,D., Griffin,J., Mouanoutoua,M., Nguyen,D., Tan,R., Rose,M., Warren,B., Ton,B., Kastury,K., Borillo,C., Carpio,T., Policky,J., Suzuki,G., Argentine,C., Shah,S., Nobrega,A., Murry,L., Turner,C., Krikorian,S., Elder,L. and Hanson,D.  
 Arabidopsis thaliana Gene Expression MicroArray  
 Unpublished (1999)  
 Contact: David Smoller, Ph.D.  
 Genome Systems, Inc., a wholly owned subsidiary of Incyte Pharmaceuticals, Inc.  
 4633 World Parkway Circle, St. Louis, MO 63134, USA  
 Tel: 877-577-2733  
 Fax: 314-427-3324  
 Email: service@genomesystems.com.

## FEATURES

## source

1. 437  
 Location/Qualifiers  
 /organism="Arabidopsis thaliana"  
 /cultivar="Columbia Col-0"  
 /db\_xref="taxon:3702"  
 /clone="701669602"  
 /clone\_lib="A. thaliana, Columbia Col-0, rosette-1"  
 /tissue\_type="rosette"  
 /dev\_stage="4 - 7 weeks"

/note="Vector: pSPORT; Site\_1: NotI; Site\_2: SalI; cDNA library was derived from untreated rosette tissue from Arabidopsis thaliana, Columbia Col-0, at 4 - 7 weeks. Plants were grown in 1:1:1 peat moss/vermiculite/perlite soil at 22 deg. C +/- 3 deg. C under constant light, and watered with fertilizer. cDNA synthesis was initiated using a NotI-oligo(dT) primer. Double-stranded cDNA was blunted, ligated to SalI adaptors, digested with NotI, size-selected, and cloned into the NotI and SalI sites of the pSPORT vector."

BASE COUNT 155 a 87 c 72 g 88 t 35 others

## ORIGIN

Query Match 3.1%; Score 11; DB 9; Length 437;  
 Best Local Similarity 100.0%; Pred. No. 27;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 238 ACNGTNATNCA 248

Db 54 ACNGTNATNCA 64

## RESULT 13

## LOCUS

AI562717 TENS2704 T. cruzi epimastigote normalized cDNA Library Trypanosoma  
 cruzi cDNA clone 2704 5', mRNA sequence.

ACCESSION AI562717.1 GI:4514062

VERSION AI562717

KEYWORDS EST.

SOURCE Trypanosoma cruzi.

ORGANISM Trypanosoma cruzi

REFERENCE Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

AUTHORS Trypanosoma; Schizotrypanum.

TITLE 1 (bases 1 to 530)

Verdun,R.E., Di Paolo,N.C., Urmenyi,T.P., Rondinelli,E., Frasch

,A.C.C. and Sanchez,D.O.

Gene discovery through expressed sequence tag sequencing in

trypanosoma cruzi

JOURNAL Infect. Immun. 66 (11), 5393-5398 (1998)

MEDLINE 99003155

COMMENT Contact: Sanchez D.O.

Instituto de Investigaciones Biotecnologicas (Univ. Nac. de Gral

San Martin)

Av. Gral Paz entre Albarillos y Constituyentes, INTI edificio 24

cp(1650) San Martin, Prov. de Bs As. Argentina

Tel: (54-1)752-9639 or (54-1)752-0021

Fax: (54-1)752-0021 or (54-1)752-9639

Email: dsanchez@inti.gov.ar

Seq primer: T7.

## FEATURES

## Location/Qualifiers

1. 530  
 /organism="Trypanosoma cruzi"  
 /strain="Cl-Brenner"  
 /db\_xref="taxon:5693"  
 /clone="2704"  
 /clone\_lib="T. cruzi epimastigote normalized cDNA Library"  
 /cell\_type="epimastigote"  
 /note="cDNA library constructed with oligo dT primed  
 epimastigote mRNA and cloned in pC7t318D phagemid with  
 modified polylinker (PHARMACIA)"

BASE COUNT 110 a 143 c 131 g 115 t 31 others

## ORIGIN

Query Match 3.1%; Score 11; DB 9; Length 530;  
 Best Local Similarity 100.0%; Pred. No. 28;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 69 NACNACNACNT 79

Db 394 NACNACNACNT 404

## RESULT 14

## LOCUS

FR0043054 Fugu rubripes GSS sequence, clone 158G14aC4, genomic survey  
 sequence.

ACCESSION AL130546

VERSION AL130546.1 GI:6112492

KEYWORDS GSS; genome survey sequence.

SOURCE Takifugu rubripes

ORGANISM Takifugu rubripes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;

Tetraodontidae; Takifugu.

REFERENCE 1 (bases 1 to 604)

AUTHORS Elgar,G., Clark,M.S., Smith,S., Meek,S., Warner,S., Edwards,Y.J.K.,

Umraniya,Y., Williams,G. and Brenner,S.

TITLE Direct Submission

JOURNAL Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource

Centre, Hinxton, Cambridge, CB10 1SB, UK Email:

biohelp@hmp.mrc.ac.uk

COMMENT Vector: pBluescript II KS

V-type: phagemid

PRIMER: KS

DESCR:

One pass dye-terminator sequencing of cosmid cloned genomic

sequence.

FEATURES

source

1. 604

Location/Qualifiers

/organism="Takifugu rubripes"

/db\_xref="taxon:31033"

/clone="158G14aC4"

/clone\_lib="cosmid 158G14"

BASE COUNT 134 a 115 c 103 g 170 t

ORIGIN

Query Match 3.1%; Score 11; DB 17; Length 604;

Best Local Similarity 100.0%; Pred. No. 30;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 89 CNACNACNCGCN 99

Db 341 CNACNACNCGCN 331

## RESULT 15

## LOCUS

BJ503599

LOCUS

DEFINITION

BJ503599 MF01FSA cDNA Oryzias latipes CDNA clone MF01FSA004A17 3',

mrna sequence.

ACCESSION

BJ503599

```

VERSION      BJ503599.1  GI:22155561
KEYWORDS     EST.
SOURCE       Japanese medaka.
ORGANISM     Oryzias latipes
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
              Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
              Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE    1 (bases 1 to 613)
AUTHORS      Kohara,Y., Shin,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
TITLE        Medaka EST Project in Takeda's lab
JOURNAL      Unpublished (2001)
COMMENT      Contact: Tadasu Shin-i
              Center For Genetic Resource Information
              National Institute of Genetics
              1111 Yata, Mishima, Shizuoka 411-8540, Japan
              Tel: 81-559-81-6856
              Fax: 81-559-81-6855
              Email: tshin@genes.nig.ac.jp.
              Location/Qualifiers
                1..613
                /organism="Oryzias latipes"
                /strain="d-rR"
                /db_xref="taxon:8090"
                /clone="MF01FSN004A17"
                /clone_lib="MF01FSA CDNA"
                /sex="mixture of female and male"
                /tissue_type="whole embryo"
                /dev_stage="fry stage 40"
BASE COUNT  159 a 123 c 105 g 206 t 20 others
ORIGIN
Query Match      3.1%; Score 11; DB 13; Length 613;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0
Qy 13 ACNATNATGCA 23
Db 279 ACNATNATGCA 289

```

Search completed: November 7, 2002, 04:58:37  
Job time : 2222 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run'on: November 6, 2002, 23:35:35 ; Search time 301 Seconds  
(without alignments)  
2626.084 Million cell updates/sec

Title: US09897438BK-2

Perfect score: 351

Sequence: 1 GARCARTGYGACNATNAT.....AYGARGCNTGYTGGCNYTN 351

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_101002:\*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*

2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*

3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*

4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*

5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*

6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*

7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*

8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*

9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*

10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*

11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*

12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*

13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*

14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*

15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*

16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*

17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*

18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*

19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*

20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*

21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*

22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:\*

23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*

24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	249.4	71.1	351	ABL40165	Mouse reelin prote
2	249.4	71.1	2745	AAD22754	Mus musculus trunc
3	249.4	71.1	11673	ABA92603	Mouse reelin encod
4	249.4	71.1	11673	ABI99284	Mouse ischaemic co
5	233.8	66.6	11580	ABN96939	Gene #3437 used to
6	233.8	66.6	11580	ABA92604	Human reelin encod
7	233.8	66.6	11632	AA589484	DNA encoding novel
8	193.8	55.2	2025	AAD22779	Xenopus laevis int
9	193.8	55.2	2274	AAD22753	Xenopus laevis tru

C 10	53.4	15.2	413	22	AAK07720	Human brain expres
C 11	53.4	15.2	413	22	AAK33553	Human bone marrow
C 12	53.4	15.2	413	24	ABS08422	Human genome-deriv
C 13	44.8	12.8	87	22	AAK20384	Human brain expres
C 14	44.8	12.8	87	22	AAK46490	Human bone marrow
C 15	38.4	10.9	4590	22	ABS20836	Human genome-deriv
C 16	38.4	10.9	4590	22	AAH24065	Yeast AOD9604-asso
C 17	37.8	10.8	6741	21	AAAI0595	Gene encoding a su
C 18	36.2	10.3	2949	21	AGD00024	Degenerate DNA enc
C 19	36.2	10.3	2949	24	AAK37455	Human zfstaz2 degen
C 20	36	10.3	759	24	AAK38781	Mouse zlmada24 dege
C 21	36	10.3	14041	22	AAH48024	Internal control B
C 22	36	10.3	14041	22	AAH48024	Internal control B
C 23	35.8	10.2	10732	21	AAAI0594	Gene encoding a su
C 24	35.6	10.1	371	22	AAI91689	Human polynucleoti
C 25	35.6	10.1	117574	24	AAI45288	Human KCNB1 gene.
C 26	35.4	10.1	485	22	ABA58007	Human foetal liver
C 27	35.4	10.1	485	22	AAK06081	Human brain expres
C 28	35.4	10.1	485	22	AAK31731	Human bone marrow
C 29	35.4	10.1	485	22	AAI37607	Probe #6293 used t
C 30	35.4	10.1	485	24	ABS06487	Human genome-deriv
C 31	35.4	10.1	1075	20	AAK84335	Stealth virus nucl
C 32	35.2	10.0	467	22	AAK82369	Human immune/haema
C 33	35.2	10.0	696	24	AAK22922	Human soluble IL-2
C 34	35.2	10.0	984	16	AAQ94336	Degenerate Alterom
C 35	35.2	10.0	3630	23	ABL27020	Drosophila melanog
C 36	35	10.0	7737	23	ABL02204	Drosophila melanog
C 37	34.8	9.9	660	19	AAV31481	Human interleukin-
C 38	34.8	9.9	660	20	AAZ40421	Human interleukin
C 39	34.8	9.9	915	22	AAH53398	S. epidermidis ope
C 40	34.8	9.9	1210	22	AAH54976	S. epidermidis gen
C 41	34.8	9.9	1905	24	AAK98781	Degenerate DNA of
C 42	34.8	9.9	3446	22	AAH54934	S. epidermidis gen
C 43	34.6	9.9	263	14	AAK61171	Human brain expres
C 44	34.6	9.9	581	22	AAH43479	(R)-2-octanol dehy
C 45	34.6	9.9	593	22	ABA61216	Human foetal liver

ALIGNMENTS

RESULT 1

ABL40165

ID ABL40165 standard; DNA; 351 BP.

AC ABL40165;

XX

XX

DT 21-MAY-2002 (first entry)

DE Mouse reelin protein CR-50 epitope region encoding DNA SEQ ID NO:1.

DE Mouse reelin protein CR-50 epitope region; elucidation; neuron;

KW Mouse; reelin protein CR-50 epitope region; elucidation; neuron;

KW cerebral disturbance; reelin protein; neuroprotective; gene; ds.

XX

OS Mus musculus.

XX

XX

Key Location/Qualifiers

CDS

FT 1..351

FT /\*tag= a

FT /product= "reelin protein CR-50 epitope region"

FT /note= "no start or stop codons given"

XX

PN JP2002017361-A.

XX

XX

PD 22-JAN-2002.

XX

XX

PF 04-JUL-2000; 2000JP-0202801.

XX

XX

PR 04-JUL-2000; 2000JP-0202801.

XX

PA (RIKA ) RIKAGAKU KENKYUSHO.

DR WPI; 2002-221707/28.  
 XX P-PSDB; ABB06244.  
 PT Reelin protein CR-50 epitope region, useful for diagnosis and treatment  
 PT of cerebral disturbance  
 XX Claim 5; Page 11; 16pp; Japanese.  
 PS  
 CC The present sequence encodes the mouse reelin protein CR-50 epitope  
 CC region, which contains the CR-50 antibody recognition site and is free  
 CC from F-spondin domains and repetitive sites. ALSO described are: (1) an  
 CC expression vector comprising a polynucleotide encoding a reelin protein  
 CC epitope region; (2) host cells with transfected the expression vector;  
 CC (3) polypeptides prepared by culture of the host cells; and (4)  
 CC polynucleotides comprising the 351 base sequence given in ABL40165 which  
 CC encodes the 117 amino acid sequence given in ABB06244; and (5) use of  
 CC the polynucleotide for diagnosis and/or treatment of diseases caused by  
 CC abnormal positioning of neural cells, and stimulation of association of  
 CC reelin protein. The mouse reelin protein CR-50 epitope region has  
 CC neuroprotective activity, and can be used in the diagnosis and treatment  
 CC of cerebral disturbance due to an abnormal reelin gene and positioning  
 CC of neurons.  
 XX  
 SQ Sequence 351 BP; 86 A; 98 C; 92 G; 75 T; 0 other;

Query Match 71.1%; Score 249.4; DB 24; Length 351;  
 Best Local Similarity 57.7%; Pred. No. 1.6e-64;  
 Matches 202; Conservative 79; Mismatches 69; Indels 0; Gaps 0;  
 QY 1 GARCARTGYGGNACNATNATGAYGNAAYCGNGTNCNTTYTGYGARCCNTAYGGNCCN 60  
 Db 1 GAGCAGTGTGGCACCACATCATGCTGCTACCTCTGTGAGCGGTACGGCCCT 60  
 QY 61 MNGARYTNACNACNACNTGYTNAAYACNACNACNACNACNACNACNACNACNACN 120  
 Db 61 CGAGAGTGTGGCACCACATCATGCTGCTACCTCTGTGAGCGGTACGGCCCT 120  
 QY 121 GGNWSNGNWSNTGYMGNTTYWSNAYGNAAYCGNACNACNACNACNACNACNACN 180  
 Db 121 GGGTCAGGATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180  
 QY 181 AAYAAACNCGNGAYTGGATNCARYTNARARATNMGNCNACNACNACNACNACNACN 240  
 Db 181 AACAAATACCGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240  
 QY 241 GTATNCAYATNTYNTAYTNCNCGARGCNAAGCNAAGCNAAGCNAAGCNAAGCNA 300  
 Db 241 GTATNCAYATNTYNTAYTNCNCGARGCNAAGCNAAGCNAAGCNAAGCNAAGCNA 300  
 QY 301 AARCAGAYWSNTYNTMGNTNGNGTNGGARGTNTAYGARGCNTGYTGGGCNT 350  
 Db 301 AAACAGACAGCCTGCGAGTGGGTGAGGTGTATGAGGCGCTGCTGGGCCCT 350

RESULT 2  
 AAD22754  
 ID AAD22754 standard; cDNA; 2745 BP.  
 XX  
 AC AAD22754;  
 XX  
 DT 26-FEB-2002 (first entry)  
 XX  
 DE Mus musculus truncated reelin cDNA.  
 XX  
 KW Mouse; reelin; F-spondin domain; CR-50 epitope; gene therapy; agyria;  
 KW polymicrogyria; ectopic gray matter; ss.  
 XX  
 OS Mus musculus.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 283..2052  
 FT /tag= a  
 FT /product= "Mouse truncated reelin protein"

FT sig\_peptide 283..363  
 FT /tag= b  
 FT mat\_peptide 364..2049  
 FT /tag= c  
 FT /product= "Mature truncated reelin protein"  
 FT misc\_feature 284..849  
 FT /tag= d  
 FT /note= "Encodes F-spondin domain"  
 FT misc\_feature 970..1320  
 FT /tag= e  
 FT /note= "Encodes CR-50 epitope region"  
 XX  
 EP1149844-A2.  
 PD 31-OCT-2001.  
 XX  
 PF 11-APR-2001; 2001EP-0303411.  
 XX  
 PR 11-APR-2000; 2000JP-0109954.  
 XX  
 PA (RIKE ) RIKEN KK.  
 XX  
 PI Mikoshiba K, Tabata H, Nakajima K;  
 XX  
 DR WPI; 2002-019320/03.  
 DR P-PSDB; AAE13606.  
 XX  
 PT Novel truncated Reelin protein containing F-spondin domain and CR-50  
 PT recognition site of Reelin protein, but not having Reelin repeat site,  
 PT useful to treat diseases including agyria due to abnormal neuron  
 PT alignment  
 XX  
 PS Claim 10; Page 20-26; 47pp; English.  
 XX  
 CC The invention relates to a truncated Reelin protein comprising a  
 CC F-spondin domain and a CR-50 recognition site but no reelin protein  
 CC repeat site. Reelin is an essential molecule in developing a normal  
 CC laminated structure of cerebrum. The truncated reelin protein and its  
 CC DNA are useful for treating diseases including agyria, polymicrogyria,  
 CC and ectopic gray matter due to abnormal neuronal alignment. Truncated  
 CC reelin protein DNA is useful in gene therapy. The present sequence is  
 CC a cDNA encoding Mus musculus truncated reelin protein.  
 XX  
 SQ Sequence 2745 BP; 661 A; 716 C; 714 G; 654 T; 0 other;

Query Match 71.1%; Score 249.4; DB 24; Length 2745;  
 Best Local Similarity 57.7%; Pred. No. 7.7e-64;  
 Matches 202; Conservative 79; Mismatches 69; Indels 0; Gaps 0;  
 QY 1 GARCARTGYGGNACNATNATGAYGNAAYCGNGTNCNTTYTGYGARCCNTAYGGNCCN 60  
 Db 970 GAGCAGTGTGGCACCACATCATGCTGCTACCTCTGTGAGCGGTACGGCCCT 1029  
 QY 61 MNGARYTNACNACNACNTGYTNAAYACNACNACNACNACNACNACNACNACNACN 120  
 Db 1030 CGAGAGTGTGGCACCACATCATGCTGCTACCTCTGTGAGCGGTACGGCCCT 1089  
 QY 121 GGNWSNGNWSNTGYMGNTTYWSNAYGNAAYCGNACNACNACNACNACNACNACN 180  
 Db 1090 GGGTCAGGATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1149  
 QY 181 AAYAAACNCGNGAYTGGATNCARYTNARARATNMGNCNACNACNACNACNACNACN 240  
 Db 1150 AACAAATACCGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1209  
 QY 241 GTATNCAYATNTYNTAYTNCNCGARGCNAAGCNAAGCNAAGCNAAGCNAAGCNA 300  
 Db 1210 GTATNCAYATNTYNTAYTNCNCGARGCNAAGCNAAGCNAAGCNAAGCNAAGCNA 1269  
 QY 301 AARCAGAYWSNTYNTMGNTNGNGTNGGARGTNTAYGARGCNTGYTGGGCNT 350  
 Db 1270 AAACAGACAGCCTGCGAGTGGGTGAGGTGTATGAGGCGCTGCTGGGCCCT 1319



CC drugs or therapeutics for ischaemic diseases. AB199913 and AB199914  
CC represent PCR primers for a mouse ischaemic condition related sequence,  
CC which are used in the exemplification of the present invention.  
XX  
SQ Sequence 11673 BP; 2831 A; 2985 C; 2985 G; 2872 T; 0 other;  
Query Match 71.1%; Score 249.4; DB 24; Length 11673;  
Best Local Similarity 57.7%; Pred. No. 2.3e-63;  
Matches 202; Conservative 79; Mismatches 69; Indels 0; Gaps 0;  
QY 1 GARCARTYGGNACNATNATGCAATGAGGNAAYGCNATNATNTTYTGCGARCCNATYGGNCCN 60  
DB 970 GAGCAGTGTGGCACCACATCATGCGCAATGCTGCTCCTCTCTGAGCCGTACGCGCT 1029  
QY 61 MNGARTYNACNACNACNATGTYTNAAYACNACNACNCCNWSNNGTNYTNCARTTYWSNATN 120  
DB 1030 CGAGCTGACACACATGCTGCAACACACACAGCATGTGCTCCAGTTTCCATT 1089  
QY 121 GGNWSNGNWSNTGYMGNTTYSNTAYWSNGAYCCNNSNATNACNATNWSNTAYGCGNAAR 180  
DB 1090 GGGTCAGGATCATGCTGATTTAGTTACTCTGACCCAGCATCACTGTGTCATACGCCAAG 1149  
QY 181 AAYAAACNCGNGAYTGGATNCARTYNGARARATNNGNCCNWSNAAAYGTNWSNACN 240  
DB 1150 AACAAATACCGCTGATGCTGATTCAGCTGGAGAAATAGAGCCCTTCCAATGTGAGCACA 1209  
QY 241 GTNATNCAYATNTYNTAYTNCNGARGCNAARGNGARWSNNGTNCARTTYCARTGG 300  
DB 1210 GTATCATCATCTCTGCTACCTCCCGAGAGGAGGAGGAGCGGTGCGAGTTCAGTGG 1269  
QY 301 AARCAGAYWSNTYMGNTNGNGARGTNTAYGARGCNTGYTGGGCGNYT 350  
DB 1270 AACAGGACAGCCTGCGAGTGGTGGTGTATGAGGCGCTGCTGGGCGCT 1319  
RESULT 5  
ABN96939  
ID ABN96939 standard; DNA; 11580 BP.  
XX  
AC ABN96939;  
XX  
DT 13-AUG-2002 (first entry)  
XX  
DE Gene #3437 used to diagnose liver cancer.  
XX  
KW Gene: liver cancer; ds; hepatocellular carcinoma; hepatotropic;  
KW metastatic liver tumor; cytostatic; expression profile; disease state;  
KW disease progression; drug toxicity; drug efficacy; drug metabolism.  
XX  
OS Homo sapiens.  
XX  
PN WO200229103-A2.  
XX  
PD 11-APR-2002.  
XX  
PF 02-OCT-2001; 2001WO-US30589.  
XX  
PR 02-OCT-2000; 2000US-237054P.  
XX  
PA (GENE-) GENE LOGIC INC.  
XX  
PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;  
XX WPI; 2002-426119/45.  
XX  
PT Diagnosing and detecting the progression of liver cancer,  
PT hepatocellular carcinoma or metastatic liver tumor in a patient,  
PT involves detecting the level of expression of two or more genes in a  
PT liver tissue sample -  
XX  
PS Claim 1; SEQ ID NO 3437; 298pp; English.  
XX  
CC The invention relates to a novel method for diagnosing and detecting the

CC progression of liver cancer, hepatocellular carcinoma or metastatic liver  
CC tumour in a patient, and differentiating metastatic liver cancer from  
CC hepatocellular carcinoma in a patient, involving detecting the level of  
CC expression of two or more genes represented in ABN93503-ABN97455 in a  
CC tissue sample. The method of the invention has hepatotropic, and  
CC cytotatic activity. The method is useful for diagnosing and detecting  
CC the progression of liver cancer, hepatocellular carcinoma and metastatic  
CC liver carcinoma in a patient. The method is useful for identifying  
CC expression profiles which serve as useful diagnostic markers as well as  
CC markers that can be used to monitor disease states, disease progression,  
CC drug toxicity, drug efficacy and drug metabolism.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 11580 BP; 3014 A; 2696 C; 2753 G; 3116 T; 1 other;  
Query Match 66.6%; Score 233.8; DB 24; Length 11580;  
Best Local Similarity 56.0%; Pred. No. 1.1e-58;  
Matches 196; Conservative 74; Mismatches 80; Indels 0; Gaps 0;  
QY 1 GARCARTYGGNACNATNATGCAATGAGGNAAYGCNATNATNTTYTGCGARCCNATYGGNCCN 60  
DB 860 GAACAGTGTGGCGCGATTTATGATGCAATGCCGTCACTCTCTGGAACCATATGCCCA 919  
QY 61 MNGARTYNACNACNATGTYTNAAYACNACNACNCCNWSNNGTNYTNCARTTYWSNATN 120  
DB 920 CGAGAACTGATACACAGCGCTTAATACACACACAGCTTCTGCTCCCAATTTCCATT 979  
QY 121 GGNWSNGNWSNTGYMGNTTYSNTAYWSNGAYCCNNSNATNACNATNWSNTAYGCGNAAR 180  
DB 980 GGGTCAGGTTTCATGCTGCTTTAGTTATTACAGCCCGAGCATCATCGTTATATGCCAAG 1039  
QY 181 AAYAAACNCGNGAYTGGATNCARTYNGARARATNNGNCCNWSNAAAYGTNWSNACN 240  
DB 1040 AATAACTCTCGGAGCTGATTCAGCTAGAGAAATAGAGCCCTTCCAATGTGAGCACA 1099  
QY 241 GTNATNCAYATNTYNTAYTNCNGARGCNAARGNGARWSNNGTNCARTTYCARTGG 300  
DB 1100 ATCATCATCATCTCTCTACCTTCTGAGGAGCGCCAAAGGAGAGATGTCCAAATTCAGTGG 1159  
QY 301 AARCAGAYWSNTYMGNTNGNGARGTNTAYGARGCNTGYTGGGCGNYT 350  
DB 1160 AAGCAGGAAATCTTCGTGTAGGTGAAGTGTATGAGGCGCTGCTGGGCGCTT 1209  
RESULT 6  
ABA92604  
ID ABA92604 standard; cDNA; 11580 BP.  
XX  
AC ABA92604;  
XX  
DT 21-MAR-2002 (first entry)  
XX  
DE Human reelin encoding cDNA SEQ ID NO:4.  
XX  
DE Human reelin; low density lipoprotein receptor; LDLR; neuroprotective;  
KW extracellular glycoprotein; nootropic; antipapenic; Alzheimer's disease;  
KW neurodegenerative disorder; neuronal regeneration; cognitive function;  
KW lipid metabolism disease; memory; developmental disorder; gene; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 176..10558  
FT /tag= a  
FT /product= "human reelin"  
XX  
PN US6323177-B1.  
XX  
PD 27-NOV-2001.  
XX  
PF 16-JUN-1999; 99US-0334220.



XX 16-JUN-1999; 99US-0334220.  
PR (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.  
PA Curran T, D'Arcangelo G;  
XX WPI; 2002-096596/13.  
DR P-PSDB; ABB05007.  
XX Novel composition useful for screening compounds that modulate Reelin  
PT binding to low density lipoprotein receptor, comprising an isolated  
PT Reelin polypeptide and low density lipoprotein receptor -  
XX Example 1; Column 75-84; 45pp; English.  
XX The present invention describes a composition (I) comprising an  
CC isolated reelin protein (II) bound to an isolated low density lipoprotein  
CC receptor (LDLR) (III). (II) is an extracellular glycoprotein of  
CC approximately 385 kDa containing a small region of similarity with  
CC R-spondin at the N terminus, a stretch of positively charged amino  
CC acids at the C terminus, and a series of eight internal repeats of  
CC 350-390 amino acids, each repeat containing two related sub-domains  
CC that flank a pattern of conserved cysteine residues known as an  
CC epidermal growth factor (EGF)-like motif. (I) has neuroprotective,  
CC neurotropic and antilipemic activities, and can be used as a modulator  
CC of reelin-LDLR interaction. (I) is useful in screen for compounds that  
CC modulate reelin binding to an LDLR, in an assay system, where the assay  
CC system comprises a microplate array and an automated robotic  
CC microprocessor controlled system for adding and removing reagents to  
CC the microplate array. The compounds identified by the above screening  
CC method are useful as therapeutic agents to provide or alleviate a  
CC diverse spectrum of diseases including neurodegenerative disorders such  
CC as Alzheimer's disease, to facilitate neuronal regeneration after  
CC injury, to prevent or alleviate lipid metabolism diseases, to enhance  
CC cognitive functions and memory or to ameliorate other developmental  
CC disorders. The present sequence encodes human reelin, which is used in  
CC the exemplification of the present invention.  
XX Sequence 11580 BP; 3014 A; 2696 C; 2753 G; 3116 T; 1 other;  
Query Match 66.6%; Score 233.8; DB 24; Length 11580;  
Best Local Similarity 56.0%; Pred. No. 1.1e-58;  
Matches 196; Conservative 74; Mismatches 80; Indels 0; Gaps 0;  
QY 1 GARCARTGYGNNACNATNATGCGAYGNAAYCGNGTACNATNTTYTGARCCNTAYGGNCCN 60  
DB 860 GAACAGTGTGGCGGATTTATGCGAATGCGCGTACCTTCTGTGAACCATATGGCCCA 919  
QY 61 MGNARVTNACNACNACNTGYTTNAAYACNACNACNCSNGNTNYTNCARTTYWSNATN 120  
DB 920 CGAGAACTGATTACCAAGCCCTTAATACAAACACAGCTTCTGTCTCCAAATTTCCATT 979  
QY 121 GGNWSNGNNSNTGYMGNTTYWSNTAYWSNGAYCCNWSNATNACNCTGTTGCTCCCAATTTCCATT 180  
DB 980 GGGTCAGTTTCATGTCGCTTGTAGTTATTCAGACCCAGCATCATCGTGTATGCGCAAG 1039  
QY 181 AAYAAACNCGNAYTGGATNCAATNCAARAATNMGNCNCCNWSNAYATGNWSNACN 240  
DB 1040 AATAACTCTCGGACTGGATTCAGCTAGAGAAAATTAGAGCCCTTCCAAATGTGACGACA 1099  
QY 241 GTNATNCAYATNTYNTAYTNCNCGARGARGCNAARGNGGARGWSNNTNCAATTYCARTGG 300  
DB 1100 ATCATCATATCTCTACCTTCTGAGAGCCGCAAGGGGAGAAATGTCCAAATTTCACTGG 1159  
QY 301 AACRAGAYNSNTNMGNGTNGNGARGTNTAYGARGCNTGYTGGGCGNYT 350  
DB 1160 AAGCAGGAAAATCTTCGTGTAGGTGAAGTGTATGAGCCCTGCTGGGCGCTT 1209  
RESULT 7  
AAS89484  
ID AAS89484 standard; cDNA; 11632 BP.

XX AAS89484;  
XX 13-FEB-2002 (first entry)  
XX DNA encoding novel human diagnostic protein #25288.  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX Homo sapiens.  
XX WO200175067-A2.  
XX 11-OCT-2001.  
XX 30-MAR-2001; 2001WO-0508631.  
XX 31-MAR-2000; 2000US-0540217.  
XX 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HYSEQ INC.  
XX Drmanac RT, Liu C, Tang YT;  
XX WPI: 2001-639362/73.  
XX P-PSDB; ABG25297.  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX Claim 1; SEQ ID NO 25288; 103pp; English.  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX Sequence 11632 BP; 3028 A; 2716 C; 2770 G; 3117 T; 1 other;  
Query Match 66.6%; Score 233.8; DB 23; Length 11632;  
Best Local Similarity 56.0%; Pred. No. 1.1e-58;  
Matches 196; Conservative 74; Mismatches 80; Indels 0; Gaps 0;  
QY 1 GARCARTGYGNNACNATNATGCGAYGNAAYCGNGTACNATNTTYTGARCCNTAYGGNCCN 60  
DB 860 GAACAGTGTGGCGGATTTATGCGAATGCGCGTACCTTCTGTGAACCATATGGCCCA 919  
QY 61 MGNARVTNACNACNACNTGYTTNAAYACNACNACNCSNGNTNYTNCARTTYWSNATN 120  
DB 920 CGAGAACTGATTACCAAGCCCTTAATACAAACACAGCTTCTGTCTCCCAATTTCCATT 979  
QY 121 GGNWSNGNNSNTGYMGNTTYWSNTAYWSNGAYCCNWSNATNACNCTGTTGCTCCCAATTTCCATT 180  
ID AAS89484 standard; cDNA; 11632 BP.



```

PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains -
XX
XX Example 4; SEQ ID NO: 7711; 650pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system,
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is one of the probes of the
XX invention.
XX
XX Sequence 413 BP; 134 A; 86 C; 67 G; 126 T; 0 other;
XX
XX Query Match 15.2%; Score 53.4; DB 22; Length 413;
XX Best Local Similarity 58.5%; Pred. No. 9.9e-06;
XX Matches 48; Conservative 14; Mismatches 20; Indels 0; Gaps 0;
XX
XX QY 1 GARCARTGYGNCNACNATNATGTCAYGGNAAAGCNGTGNACNTTTCYTGARCCNTAYGGNCCN 60
XX |||::|||::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 323 GAACAGTGTGCGCGATATATGATCGGAATGCCGTACCTTCTGTGAACCATATGGCCCA 264
XX
XX QY 61 MNGARYTNACNACNACNTGYY 82
XX : |||::||| ||| ||| ||| |||
XX Db 263 CGAGAACTGGTAAGTATGTGCT 242
XX
XX RESULT 11
XX AAK33553/c
XX ID AAK33553 standard; DNA; 413 BP.
XX
XX AC AAK33553;
XX
XX XX
XX DT 06-NOV-2001 (first entry)
XX
XX DE Human bone marrow expressed single exon probe SEQ ID NO: 8110.
XX
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200157276-A2.
XX
XX XX 09-AUG-2001.
XX
XX XX 30-JAN-2001; 2001WO-US00668.
XX
XX XX 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 03-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains -
XX
XX Example 4; SEQ ID NO: 7711; 650pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system,
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is one of the probes of the
XX invention.
XX
XX Sequence 413 BP; 134 A; 86 C; 67 G; 126 T; 0 other;
XX
XX Query Match 15.2%; Score 53.4; DB 22; Length 413;
XX Best Local Similarity 58.5%; Pred. No. 9.9e-06;
XX Matches 48; Conservative 14; Mismatches 20; Indels 0; Gaps 0;
XX
XX QY 1 GARCARTGYGNCNACNATNATGTCAYGGNAAAGCNGTGNACNTTTCYTGARCCNTAYGGNCCN 60
XX |||::|||::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 323 GAACAGTGTGCGCGATATATGATCGGAATGCCGTACCTTCTGTGAACCATATGGCCCA 264
XX
XX QY 61 MNGARYTNACNACNACNTGYY 82
XX : |||::||| ||| ||| ||| |||
XX Db 263 CGAGAACTGGTAAGTATGTGCT 242
XX
XX RESULT 11
XX AAK33553/c
XX ID AAK33553 standard; DNA; 413 BP.
XX
XX AC AAK33553;
XX
XX XX
XX DT 06-NOV-2001 (first entry)
XX
XX DE Human bone marrow expressed single exon probe SEQ ID NO: 8110.
XX
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200157276-A2.
XX
XX XX 09-AUG-2001.
XX
XX XX 30-JAN-2001; 2001WO-US00668.
XX
XX XX 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 03-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains -
XX
XX Example 4; SEQ ID NO: 7711; 650pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system,
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is one of the probes of the
XX invention.
XX
XX Sequence 413 BP; 134 A; 86 C; 67 G; 126 T; 0 other;
XX
XX Query Match 15.2%; Score 53.4; DB 22; Length 413;
XX Best Local Similarity 58.5%; Pred. No. 9.9e-06;
XX Matches 48; Conservative 14; Mismatches 20; Indels 0; Gaps 0;
XX
XX QY 1 GARCARTGYGNCNACNATNATGTCAYGGNAAAGCNGTGNACNTTTCYTGARCCNTAYGGNCCN 60
XX |||::|||::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 323 GAACAGTGTGCGCGATATATGATCGGAATGCCGTACCTTCTGTGAACCATATGGCCCA 264
XX
XX QY 61 MNGARYTNACNACNACNTGYY 82
XX : |||::||| ||| ||| ||| |||
XX Db 263 CGAGAACTGGTAAGTATGTGCT 242
XX
XX RESULT 11
XX AAK33553/c
XX ID AAK33553 standard; DNA; 413 BP.
XX
XX AC AAK33553;
XX
XX XX
XX DT 06-NOV-2001 (first entry)
XX
XX DE Human bone marrow expressed single exon probe SEQ ID NO: 8110.
XX
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200157276-A2.
XX
XX XX 09-AUG-2001.
XX
XX XX 30-JAN-2001; 2001WO-US00668.
XX
XX XX 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 03-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains -
XX
XX Example 4; SEQ ID NO: 7711; 650pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system,
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is one of the probes of the
XX invention.
XX
XX Sequence 413 BP; 134 A; 86 C; 67 G; 126 T; 0 other;
XX
XX Query Match 15.2%; Score 53.4; DB 22; Length 413;
XX Best Local Similarity 58.5%; Pred. No. 9.9e-06;
XX Matches 48; Conservative 14; Mismatches 20; Indels 0; Gaps 0;
XX
XX QY 1 GARCARTGYGNCNACNATNATGTCAYGGNAAAGCNGTGNACNTTTCYTGARCCNTAYGGNCCN 60
XX |||::|||::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX 323 GAACAGTGTGCGCGATATATGATCGGAATGCCGTACCTTCTGTGAACCATATGGCCCA 264
XX
XX QY 61 MNGARYTNACNACNACNTGYY 82
XX : |||::||| ||| ||| ||| |||
XX Db 263 CGAGAACTGGTAAGTATGTGCT 242
XX
XX RESULT 11
XX AAK33553/c
XX ID AAK33553 standard; DNA; 413 BP.
XX
XX AC AAK33553;
XX
XX XX
XX DT 06-NOV-2001 (first entry)
XX
XX DE Human bone marrow expressed single exon probe SEQ ID NO: 8110.
XX
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200157276-A2.
XX
XX XX 09-AUG-2001.
XX
XX XX 30-JAN-2001; 2001WO-US00668.
XX
XX XX 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 03-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains -
XX
XX Example 4; SEQ ID NO: 7711; 650pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system,
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is one of the probes of the
XX invention.
XX
XX Sequence 413 BP; 134 A; 86 C; 67 G; 126 T; 0 other;
XX
XX Query Match 15.2%; Score 53.4; DB 
```







Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	249.4	71.1	11673	4	US-09-334-220-3	Sequence 3, Appli
2	233.8	66.6	11580	4	US-09-334-220-4	Sequence 4, Appli
c 3	39.2	11.2	7218	1	US-08-232-463-14	Sequence 14, Appli
4	36.2	10.3	2949	4	US-09-412-554-3	Sequence 3, Appli
c 5	34.4	9.8	624	4	US-09-397-992A-3	Sequence 3, Appli
6	34	9.7	37948	4	US-09-251-645-11	Sequence 11, Appl
7	33.8	9.6	628	4	US-09-385-982-149	Sequence 149, App
c 8	33.4	9.5	176373	3	US-09-128-155-17	Sequence 17, Appl
9	33.2	9.5	1248	4	US-09-134-001C-2240	Sequence 2240, Ap
10	33.2	9.5	152331	3	US-09-128-155-16	Sequence 16, Appl
11	33	9.4	957	4	US-09-134-001C-543	Sequence 543, App
c 12	32.8	9.3	624	4	US-09-397-992A-6	Sequence 6, Appli
13	32.6	9.3	2057	4	US-09-106-194-11	Sequence 11, Appl
c 14	32.6	9.3	7498	3	US-08-816-693A-1	Sequence 1, Appli
c 15	32.6	9.3	7498	3	US-08-885-291-1	Sequence 1, Appli
c 16	32.6	9.3	7498	4	US-09-496-672-1	Sequence 1, Appli
17	32.4	9.2	6743	3	US-08-932-280-1	Sequence 1, Appli
18	32.4	9.2	12949	4	US-09-538-414-11	Sequence 11, Appl
c 19	32.2	9.2	2294	4	US-09-623-591-123	Sequence 123, App
c 20	32	9.1	5476	4	US-09-241-017B-147	Sequence 147, App
21	31.8	9.1	31880	4	US-09-453-702B-242	Sequence 242, App
c 22	31.8	9.1	246240	2	US-08-724-394A-20	Sequence 20, Appl
23	31.8	9.1	246240	2	US-08-724-394A-21	Sequence 21, Appl
24	31.8	9.1	246240	2	US-08-724-394A-22	Sequence 22, Appl
25	31.6	9.0	361	1	US-08-053-131-171	Sequence 171, App
c 26	31.6	9.0	361	1	US-08-096-762-171	Sequence 171, App
27	31.6	9.0	2932	2	US-08-481-337A-5	Sequence 5, Appli

```
RESULT 2
US-09-334-220-4
; Sequence 4, Application US/09334220
; Patent No. 6323177
; GENERAL INFORMATION:
; APPLICANT: St. Jude Children's Research Hospital
; APPLICANT: Curran, Thomas
; APPLICANT: D'Arcangelo, Gabriella
; TITLE OF INVENTION: INTERACTION OF REELIN WITH VERY LOW
; TITLE OF INVENTION: DENSITY LIPOPROTEIN (VLDL) RECEPTOR FOR SCREENING AND
; TITLE OF INVENTION: THERAPIES
; FILE REFERENCE: 2427/0F704
; CURRENT APPLICATION NUMBER: US/09/334,220
; CURRENT FILING DATE: 1999-06-16
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows version 3.0
; SEQ ID NO 4
; LENGTH: 11580
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-334-220-4

Query Match          66.6%; Score 233.8; DB 4; Length 11580;
Best Local Similarity 56.0%; Pred. No. 2.9e-67;
Matches 196; Conservative 74; Mismatches 80; Indels 0; Gaps 0;

QY 1 GARCARTGYGNCNATNATGCAAGGNAAYGCNCTNACNTTYTGYGARCCTAYGGNCCN 60
Db 860 GACAGTGTGGCGGATATGCAATGGCAATGCCGTACCTCTGTGAACCATATGGCCCA 919
QY 61 MGARYATNACNACNACNTGYTNAAYACNACNACNCSNGTNTNCTARTTYSNATN 120
Db 920 CGAAGCTGATACCAAGCCCTTAATACACACAGCTTCTGCTCCCAATTTCCATT 979
QY 121 GGNWSNGGNWSNTGYMNTTYSNTAYWSNGAYCCNWSNATNACNCTNWSNTAYGCAAR 180
Db 980 GGGTCAGGTTTCATGTCGCTTAGTATTCAGACCCAGCATCATCGTGTATATGCCAAG 1039
QY 181 AYAAYACNCGNAYTGGATNCARYTNGARAATNMGNCNCSNNAAYGTNWSNACN 240
Db 1040 AATAACTCTCGGACTGGATGATCAGTAGAGAAATAGAGCCCTTCCCAATGTCAGACA 1099
QY 241 GTNATNCAYATNTNTAYTNCNGARGCNAARGNAGNWSNCTNCTARTTYCARTGG 300
Db 1100 ATCATCATATCTCTACCTTCTGAGGACGCCAAGGGAGAGATGTCCAATTTCACTGG 1159
QY 301 AARCAGAYWSNTNMGNTNGNGARGTNTAYGARGCNTGYTGGGNYT 350
Db 1160 AAGCAGGAAAATCTCTGTGTAGTGAAGTGTATGAAGCCCTGCTGGGCCTT 1209

RESULT 3
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)883-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
US-08-232-463-14

Query Match          11.2%; Score 39.2; DB 1; Length 7218;
Best Local Similarity 7.4%; Pred. No. 0.037;
Matches 25; Conservative 140; Mismatches 172; Indels 0; Gaps 0;

QY 1 GARCARTGYGNCNATNATGCAAGGNAAYGCNCTNACNTTYTGYGARCCTAYGGNCCN 60
Db 1448 GAAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1389
QY 61 MGARYATNACNACNACNTGYTNAAYACNACNACNCSNNGTNTNCTARTTYSNATN 120
Db 1388 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1329
QY 121 GGNWSNGGNWSNTGYMNTTYSNTAYWSNGAYCCNWSNATNACNCTNWSNTAYGCAAR 180
Db 1328 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1269
QY 181 AYAAYACNCGNAYTGGATNCARYTNGARAATNMGNCNCSNNAAYGTNWSNACN 240
Db 1268 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1209
QY 241 GTNATNCAYATNTNTAYTNCNGARGCNAARGNAGNWSNCTNCTARTTYCARTGG 300
Db 1208 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1149
QY 301 AARCAGAYWSNTNMGNTNGNGARGTNTAYGARG 337
Db 1148 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1112

RESULT 4
US-09-412-554A-3
; Sequence 3, Application US/09412554A
; Patent No. 6355788
; GENERAL INFORMATION:
; APPLICANT: Conklin, Darrell
; APPLICANT: Ellsworth, Jeff L.
; TITLE OF INVENTION: FOLLISTATIN RELATED PROTEIN 2FSTA2
; FILE REFERENCE: 98-50
; CURRENT APPLICATION NUMBER: US/09/412,554A
; CURRENT FILING DATE: 1999-10-05
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 2949
; TYPE: DNA
```



Db	325	AYTTYTCNGTGTGRTTYCTCCACNCCTCARNNSDATTGTCNCCKBAARNWSRA	2656
Qy	97	GCNWSNGTNTNCARTTYSWNTATNGGNWSNGTGYMGNTTYSWNTAYWSNGAYCCN	1567
Db	265	ADATYTCGTGNARCATYTCGNARDATNGCNAIRGTRTCNCCYTYTGTAATGYTGNC	2067
Qy	157	WSNATNACNGTNSWNTAYGCNAARAAYACNGCNGAYTGGATNCARYNGARAARATN	2167
Db	205	GSNWARNWSYTTTGTGGGNARNARRAARTYTTNCKRTGTCNGGNARRCATYGTGTGATNS	1467
Qy	217	MGNCGCCNWSNAAAGTNWSNACNGTNA TCATATNTAYTNCNGCNGARGAGCNAAR	2767
Db	145	WNRNGTYTGNAARYTTTNNARNYTTNARNSWTTCYTGRTNACTGTGCKTGTGTGRA	867
Qy	277	GGNGARWSNGTNCARTTYCATGGAARCARCAYWSNTNMGNTNGGNGGAR	3277
Db	85	ADATDATNARYTTTWARCTCNARNSWRAADATNGTNGTNSWNGCNARNARNA	357

## RESULT 6

```

US-05-231-643-11
; Sequence 11, Application US/09251645
; Patent No. 6281413
; GENERAL INFORMATION:
; APPLICANT: Kramer, Vance C.
; APPLICANT: Morgan, Michael K.
; APPLICANT: Anderson, Arne R.
; APPLICANT: Hart, Hope
; APPLICANT: Warren, Gregory W.
; APPLICANT: Dunn, Martha
; APPLICANT: Chen, Jeng S.
; TITLE OF INVENTION: NOVEL INSECTICIDAL TOXINS FROM PHOTORHABDUS LUMINESCENS
; TITLE OF INVENTION: AND NUCLEIC ACID SEQUENCES CODING THEREFOR
; FILE REFERENCE: GCG1963/A
; CURRENT APPLICATION NUMBER: US/09/251,645
; CURRENT FILING DATE: 1999-02-17
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0

```

## RESULT 5

```

? CURRENT APPLICATION NUMBER: US/09/251,645
?
? CURRENT FILING DATE: 1999-02-17
?
? NUMBER OF SEQ ID NOS: 22
?
? SOFTWARE: PatentIn Ver. 2.0
?
? SEQ ID NO 11
?
? LENGTH: 37948
?
? TYPE: DNA
?
? ORGANISM: Photorhabdus luminescens
?
? FEATURE:
?
? NAME/KEY: CDS
?
? LOCATION: (15171)..(18035)
?
? OTHER INFORMATION: orf5
?
? FEATURE:
?
? NAME/KEY: CDS
?
? LOCATION: (23768)..(31336)
?
? OTHER INFORMATION: hph2
?
? FEATURE:
?
? NAME/KEY: CDS
?
? LOCATION: (31393)..(35838)
?
? OTHER INFORMATION: orf2
?
? US-09-251-645-11

```

Query Match	9.7%;	Score 34;	DB 4;	Length 37948;
Best Local Similarity	30.9%;	Pred. No. 9.4;		
Matches 43;	Conservative 26;	Mismatches 70;	Indels 0;	Gaps
Qy 76	ACNTGYTNAAYACNACNACNCWNGTNTYTCARTTYSNATNGNWSNGCNWNTGY 135			
Db 14454	ACCATGCTCATGATCCATTTGTCGAGAAATCAAGTCTACTACCCGCAAAAGCTGGTTC 145			
Qy 136	MGNTTYSNTAYWSNGAYCCNWSNAYNMACNNGTNTYGCNARAAAYAYACNCGNGAY 195			
Db 14514	CGTCGAACCTTGTTCACTCCCTGGTTTACTGTCTAATGAAGATGAATAATGCACAGCTACT 145			
Qy 196	TGGATNCARYTNGARAARA 214			
Db 14574	GAGGTGAAGGTTAAAGAAGA 14592			

```

RESULT 7
US-09-385-982-149
; Sequence 149, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 149
; LENGTH: 628
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(628)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-149

Query Match          9.6%; Score 33.8; DB 4; Length 628;
Best Local Similarity 25.3%; Pred.No.0.25;
Matches 57; Conservative 39; Mismatches 129; Indels 0; Gaps

Qy 67 YTNACNACNACNTGYTNAAYACNACNACNCGNWSNGTYNCARTYWSNATNGNWSN 126
   :| || || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 31 TTTTNAACAGCGNCTTTTCATTTTTTATTACTCAAAAAGTTTCATTTTTTTTAAAGT 90

Qy 127 GGNWSNTGYMGNITYNSNTAYWSNGAYCCNWSNATVACNGTWSNATYCGNRAARAAY 186
   :: :| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 91 TTTCTGACTCTGNGCCTTGGGCGCTTCAACACTTTTCACAACGATTTTCTGCTCCTCGTAAGG 150

Qy 187 ACNCGNCAVYTGGAATNCARYTNGARAARTNMGNCNCCNWSNAAAYCTNWSNACNGTNTATN 246
   | || || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 151 AAAGCCGCGCTTGATCTCTANAAAGGAAAATACCAAAATTAATCATCTTCTTAAATGAAC 210

Qy 247 CAYATNVTNTAYTNCNCGARGARGCNAAAGNGGARWSNGTNCAR 291
   || :| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 211 CATTTTATTTTAGCCCCAAAAGGNAACATGTAAGAAAGCAAA 255

```

```

RESULT 8
US-09-128-155-17/c
; Sequence 17, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 176373
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature

```

[illegible]

```

Query Match          9.5%; Score 33.2; DB 4; Length 1248;
Best Local Similarity 28.4%; Pred. No. 0.75;
Matches 50; Conservative 30; Mismatches 96; Indels 0; Gaps 0;

QY      35  TNAGNTTGTGARGCCNTAYGNCNMGNGARYRYNACNACNACNTGGYYTNAAYACNACNA 94
      Db      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      95  CNGCNWSNGTWTYNCARTTYSNATNGNGNWSNGTGYMGNNTTYSNATYWSNGAYC 154
      Db      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY      869  CCTCCACCATGACGGGCATAACCCACCATACGTATATCACTATAATTTTACGTCACGTTAAA 810
      Db      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY      155  CNWSNATNACNGTWNSTAYGCNARAYAYACNGCNGAYTGGCATNCARYTNGAR 210
      Db      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY      809  CCAGATCTCTTGAGGTCCTCCAATAACGAACGTCACGTCGGGTTAATGTAAAA 754
      Db      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 10
US-09-128-1155-16
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; TITLE OF INVENTION: AND USE THEREOF
; FILE REFERENCE: 09404/052001

```



APPLICANT: Jelinek, Laura  
APPLICANT: Durnam, Diane  
APPLICANT: Blumberg, Hal  
TITLE OF INVENTION: NOVEL NUCLEAR RECEPTOR POLYPEPTIDE  
TITLE OF INVENTION: ZPPAR4  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ZymoGenetics, Inc.  
STREET: 1201 Eastlake Avenue East  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98102  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/106,194  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Leith, Debra K  
REGISTRATION NUMBER: 32,619  
REFERENCE/DOCKET NUMBER: 96-11  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-442-6674  
TELEFAX: 206-442-6678  
TELEX:  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2067 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-106-194-11

Query Match 9.3%; Score 32.6; DB 4; Length 2067;

Best Local Similarity 9.7%; Pred. No. 1.9;

Matches 29; Conservative 92; Mismatches 177; Indels 1; Gaps 1;

QY 1 GACARTGYGNCACNATNATGCAAGGNAAYGCNGTACNATTTTCYGCARCCNTAYGGCCN 60  
Db 749 GNMGNWNSNCARCAAAATGCGNSNTAYWSNTGYCCNMGNCAARMGNAAYGTYTTHG 808  
QY 61 MGNGARTNACNACNACNTGYTNAAYAC-NACNACGNCNWSNGTNYTNCARTTYWSNAT 119  
Db 809 AYGNCACNAAYMGNAAYMGNTGYCARCAATGYMGNTNCARAARTGYTNGCNYTNGGNA 868  
QY 120 NGGNWSNGNWSNGTYGNTTYWSNTAYWSNGAYCCNWSNATNACNCTNWSNTAYGCNAA 179  
Db 869 TGWSNMNGAYGCGNTNAARTTYGNGNATGWSNARAARCAARMGNGAYWSNTAYG 928  
QY 180 RAAYAYACNGCNGAYTGGATNCARYTNGARAARTNMGNCNCCNWSNAYGTTNWSNAC 239  
Db 929 CNGARTNCARAARCAATCARGCMGNTNCARGARCMGNCARCAARWSNGNGARG 988  
QY 240 NGTNCATNAYATNTAYTNCNCGARGCNAARGGNGNWSNCTNCARTTYCAT 298  
Db 989 CNGARGCNYTNCMGNGTNTAYWSNWSNATHWSNAAAYGGNTYWSNAAAYTNAAY 1047

RESULT 14

US-08-816-693A-1/c

; Sequence 1, Application US/08816693A

; Patent No. 5874241

; GENERAL INFORMATION:

; APPLICANT: Takahashi, Joseph S

; APPLICANT: Turek, Fred W

APPLICANT: Pinto, Lawrence H  
TITLE OF INVENTION: Clock Gene and Gene Product  
NUMBER OF SEQUENCES: 53  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dressler, Rocky, Milnamow & Katz  
STREET: Two Prudential Plaza, Suite 4700  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/816,693A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5874241thrup, Thomas E  
REGISTRATION NUMBER: 33,268  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-616-5400  
TELEFAX: 312-616-5460  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7498 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 389..2954  
US-08-816-693A-1

Query Match 9.3%; Score 32.6; DB 2; Length 7498;

Best Local Similarity 24.8%; Pred. No. 6.2;

Matches 62; Conservative 43; Mismatches 145; Indels 0; Gaps 0;

QY 52 TAYGGNCCNMNGARYTNACNACNACNTGYTNAAYACNACNACNWSNGNWTNCTNAR 111  
Db 1156 TCTGACAGTAGTACAAAACAACTCTATCTTATAAGAGGCTATGTGCGTCTAT 1097  
QY 112 TTWSNATNGNWSNGNWSNTGYMGNTTYWSNTATWSNGAYCCNWSNATNACNCTNWSN 171  
Db 1096 AGTTCTCTCAAAACCAATGTGTGTGAAGTTGATACACTGGTTAAAGATTAAATTC 1037  
QY 172 TAYGCNAARAAYAYACNCGNGAYTGGATNCARYTNGARAARATNMGNCNCCNWSNAA 231  
Db 1036 TATAAATCTCACATATTCATAGTGGATGGCTCTCTTTGGGTCTATTTCTCGAAGCAT 977  
QY 232 GTNWSNACNCTNATNCAATNTYNTAYTNCNCGARGCNAARGGNGNWSNGTNCAR 291  
Db 976 GTGACAAACAGATTCCTAACTGATTTTGTGATTTTAACTACTCAGGGGTTAATGAGTCACT 917  
QY 292 TTYCARTGGA 301  
Db 916 TTCACGAGA 907

RESULT 15

US-08-885-291-1/c

; Sequence 1, Application US/08885291A

; Patent No. 6057125

; GENERAL INFORMATION:

; APPLICANT: Takahashi, Joseph S.

; APPLICANT: Turek, Fred W.

; APPLICANT: Pinto, Lawrence H.

; TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT

; FILE REFERENCE: 0290-5

; CURRENT APPLICATION NUMBER: US/08/885,291A

```

; CURRENT FILING DATE: 1997-06-30
; EARLIER APPLICATION NUMBER: 08/816,693
; EARLIER FILING DATE: 1997-03-13
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 7498
; TYPE: DNA
; ORGANISM: Mus musculus
US-08-885-291-1

```

```

Query Match      9.3%; Score 32.6; DB 3; Length 7498;
Best Local Similarity 24.8%; Pred. No. 6.2;
Matches 62; Conservative 43; Mismatches 145; Indels 0; Gaps 0;

QY 52 TAYGGNCCNMGNGARYTNACNACNACNTGYTNAAYACNACNACNCGNWSNGTNYTNCAR 111
Db 1156 TCTGACAGTAGCTACAAACAACTCTATCTTCATAAGAGGCCCTATGTGTGCGTTGTAT 1097

QY 112 TTYWSNATNGNWSNGNWSNTGYMNTTYWSNTAYWSNGAYCCNWSNATNACNGTNWSN 171
Db 1096 AGTTCCTTCAAAACCATTTGCTGTGAAGTTGATACACTGGTTAAAGATTTAAATTTCC 1037

QY 172 TAYGCNAARAAYAYACNGNGAYVTGGATNCARYTNGARAARATNMGNGCCNWSNAA 231
Db 1036 TATAAATCTCACATATTTCATAGGTGGATGGCTCCTTTGGGTCTATTGTTCTCGAAGCAT 977

QY 232 GTNWSNACNGTNATNCAYATNTYATYTNCCNGARGCNGARNGGNGARWSNGTNCAR 291
Db 976 GTGACAAACAGAAATCTTAACATGATTTTGTGATTTTAAGTACTCAGGGGTTAATGAGTCACT 917

QY 292 TTYCARTGGA 301
Db 916 TTCAGCAGA 907

```

Search completed: November 7, 2002, 02:19:44  
Job time : 156 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 01:45:40 ; Search time 87 Seconds  
(without alignments)  
1344.107 Million cell updates/sec

Title: US09897438BK-2  
Perfect score: 351  
Sequence: 1 GARCARTGGNACNATNAT.....AYGARGCNRTGTGGGCVN 351

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 310279 seqs, 166577418 residues  
Total number of hits satisfying chosen parameters: 620558

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	249.4	71.1	351	US-09-897-438B-1	Sequence 1, Appli
2	233.8	66.6	11580	US-09-880-107-3436	Sequence 3436, Ap
3	90.2	25.7	499	US-09-864-761-14497	Sequence 14497, A
4	78.2	22.3	214	US-09-864-761-31037	Sequence 31037, A
5	53.4	15.2	413	US-09-864-761-11808	Sequence 11808, A
6	44.8	12.8	87	US-09-864-761-28379	Sequence 28379, A
7	35.4	10.1	485	US-09-864-761-10999	Sequence 10999, A
8	35.2	10.0	696	US-09-825-561A-8	Sequence 8, Appli
9	34.8	9.9	1905	US-09-922-488-3	Sequence 3, Appli
10	34.6	9.9	593	US-09-864-761-7539	Sequence 7539, Ap
11	34.6	9.9	594	US-09-917-800A-1003	Sequence 1003, Ap
12	34.6	9.9	4068	US-09-962-436-295	Sequence 295, App
13	34.4	9.8	12932	US-09-764-847-1132	Sequence 1132, Ap
14	33.6	9.6	921	US-09-925-300-223	Sequence 223, App
15	33.4	9.5	1278	US-09-815-242-6280	Sequence 6280, Ap
16	33.4	9.5	180216	US-09-835-232-6	Sequence 6, Appli
17	33.2	9.5	560	US-09-864-761-32024	Sequence 32024, A
18	33.2	9.5	1985	US-09-864-761-15510	Sequence 15510, A
19	33.2	9.5	2379	US-10-006-867-123	Sequence 123, App

ALIGNMENTS

RESULT 1  
US-09-897-438B-1  
; Sequence 1, Application US/09897438B  
; Patent No. US20020137095A1  
; GENERAL INFORMATION:  
; APPLICANT: Mikoshiba, Katsuhiko  
; APPLICANT: Tate, Naoko  
; TITLE OF INVENTION: REELIN PROTEIN CR-50 EPI TOPE REGION  
; FILE REFERENCE: 04853-0076-00000  
; CURRENT APPLICATION NUMBER: US/09/897,438B  
; PRIOR FILING DATE: 2001-07-03  
; PRIOR APPLICATION NUMBER: JP 2000-202801  
; PRIOR FILING DATE: 2000-07-04  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 351  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-09-897-438B-1

Query Match 71.1%; Score 249.4; DB 10; Length 351;  
Best Local Similarity 57.7%; Pred. No. 1.3e-65;  
Matches 202; Conservative 79; Mismatches 69; Indels 0; Gaps 0;  
QY 1 GARCARTGGNACNATNATGTCAYGGNAAAYCGGTCNACNTTYTGYGARCCTAYGGCCN 60  
DB 1 GAGCAGTGTGGCACCATCATGTCATGCAATGCTGTACCTTCTGTGAGCCGTACGCCCT 60  
QY 61 MGNGARYTNACNACNACNTGYTNAAYACNACNACNACNACNACNACNACNACNACNACN 120  
DB 61 CGAGAGCTGACCCACCATGCTGCAACACACACATCATGCTCCTCCATGTTTCATT 120  
QY 121 GGWNSGNGNSNTGYMGNTTYTWSNTAYWSNGAYCCNWSNATNACNAGTNGNATNACN 180  
DB 121 GGGFCAGGATCATGCTGATTTAGTTACTCTGACCCAGCATGCTGTCTATACGCCAAG 180  
QY 181 AATYAACNCGNGAYTGATNCAATNCAATNCAATNCAATNCAATNCAATNCAATNCAATN 240  
DB 181 AACAAATCCGCTGATTTGATTCAGTGGAGAAAATAGAGCCCTTTCATATGTGAGACA 240  
QY 241 GTNATNCAATNTYNTAYTYNCCNGARGCNAARGNARWSNGNATNCAATTCATGCG 300

```

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 14497
; LENGTH: 499
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC000121.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.62
US-09-864-761-14497

Query Match 25.7%; Score 90.2; DB 10; Length 499;
Best Local Similarity 57.5%; Pred. No. 1.5e-17;
Matches 77; Conservative 27; Mismatches 30; Indels 0; Gaps 0;

QY 217 MGNGCNCNWSNAAATGNNACNGTNCATNTATYVTTNCNGARGAGCNAAR 276
Db 487 AGAGCCCCCTCCATCTCAGCAATCATCATCTTACCTCTCTGAGGAGCCAAA 428
QY 277 GNGARWSNGTNCARTTYCARTGCAARCAARGAYWSNYTNMNGTNGNGARGTNTAYGAR 336
Db 427 GGGGACAATGTCCAAATTCAGTGAAGCAAGGAAAAATCTTCGTAGTGAAGTGTATGAA 368
QY 337 GCNTGTGGCNYT 350
Db 367 GCCTGCTGGGCCTT 354

RESULT 4
US-09-864-761-31037/c
; Sequence 31037, Application US/09864761
; Patent No. US20020048763A1
```

```

; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3436
; LENGTH: 11580
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 U79716
US-09-880-107-3436

Query Match 66.6%; Score 233.8; DB 10; Length 11580;
Best Local Similarity 56.0%; Pred. No. 1.2e-59;
Matches 196; Conservative 74; Mismatches 80; Indels 0; Gaps 0;

QY 1 GARCARTGYGNGACNATNATGAYGNAAYGCGNGTNCACNTTYGYGARCNTAYGNGCCN 60
Db 860 GACAGGTGGCGGATATGCAATGCAATGCCGTCCCTCTGTGAACCATATGGCCCA 919
QY 61 MGNGARYTNACNACNACNTGYTNAAYACNACNACNACNACNACNACNACNACNACN 120
Db 920 CGAAGCTGATTACCAAGCCCTTAATACAAACACAGCTTCTGCTCCCAATTTCCATT 979
QY 121 GGWSNGNWSNTGYMNTTYWSNTAYWSNGAYCCNWSNATNACNNGTNCNWSNTAYG 180
Db 980 GGGTCAGTTTCATGTCGCTTTAGTTATTCAGACCCAGCATCATCTGTTATATGCAAG 1039
QY 181 AAYAAACNCGNGAYTGGATNCAATYNGARAARATNMNGCNCNWSNAAATGNNACN 240
Db 1040 AATAACTCTCGGACTGGATGATGATGAGTAAATAGAGCCCTTCCCAATGTCAGACA 1099
QY 241 GTNATNCAYATNTYATYTNCCNGARGCNAARGNGNGARWSNGTNCARTTYCARTGG 300
Db 1100 ATCATCATATCTCTACCTTCTGAGGAGCCCAAGGGAGAGATGTCCAATTTCACTGG 1159
QY 301 AACARGAYWSNTYMGNTNGNGARGTNTAYGARGCNTGYTGGCNYT 350
Db 1160 AAGCAGAAAAATCTCTGTGATGTAAGTGTATGAAGCCTGCTGGGCCTT 1209

RESULT 3
US-09-864-761-14497/c
; Sequence 14497, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
```







```
Best Local Similarity 31.4%; Pred. No. 0.47;
Matches 54; Conservative 27; Mismatches 91; Indels 0; Gaps 0;

Qy 160 ATNACNGTNSNTAYGCNNAARAAAYACNCGNCAYTGGATNCARYTNGARAARATWGN 219
Db 303 ACAAAAATTAAGACAAAGTTACTGACTTCAGTTTCCTTTGATTGATGAGTGAACA 244
Qy 220 GCNCCNNSAAYGTNWSNACNGTNCATNCAYATNTNTAYTNCNCGARGCGCNAARGN 279
Db 243 ACACATTGCGTGTGAACAATTTATTACACATCTTGGATGTCCTTAAGGGACGACGAGGA 184
Qy 280 GARWSNCTNCARTTYCARTGGAARCAARGAYWSNTNNMGNTNGNGARGTNT 331
Db 183 GGGAAATTGTTATTCACATGCAAAAGAGTGTGTGTGTCATGCCAGGGCT 132

RESULT 8
US-09-825-561A-8
: Sequence 8, Application US/09825561A
: Patent No. US20020137677A1
: GENERAL INFORMATION:
: APPLICANT: Sprecher, Cindy A.
: APPLICANT: No. US20020137677A1ak, Julia E.
: APPLICANT: West, James W.
: APPLICANT: Presnell, Scott R.
: APPLICANT: Holly, Richard D.
: APPLICANT: Nelson, Andrew J.
: TITLE OF INVENTION: SOLUBLE ZALPHAL1 CYTOKINE RECEPTORS
: FILE REFERENCE: 00-22
: CURRENT APPLICATION NUMBER: US/09/825,561A
: CURRENT FILING DATE: 2000-04-05
: PRIOR APPLICATION NUMBER: US 60/194,731
: PRIOR FILING DATE: 2000-04-05
: PRIOR APPLICATION NUMBER: US 60/222,121
: PRIOR FILING DATE: 2000-07-28
: NUMBER OF SEQ ID NOS: 86
: SOFTWARE: FastSEQ for Windows Version 3.0
: SEQ ID NO 8
: LENGTH: 696
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Degenerate polynucleotide sequence of IL-2Rgamma
: NAME/KEY: misc_feature
: LOCATION: (1)..(696)
: OTHER INFORMATION: n = A,T,C or G
US-09-825-561A-8

Query Match 10.0%; Score 35.2; DB 10; Length 696;
Best Local Similarity 12.1%; Pred. No. 0.72;
Matches 28; Conservative 68; Mismatches 136; Indels 0; Gaps 0;

Qy 114 YWSNATNGNWSNGNWSNTGYMGNTTYSNTAYWSNGAYCCNWSNATNACNGTWSNNTA 173
Db 418 YTNCAAYARTNWSNGARWSNCARTYNGARYTNAAYTGGAAAYAYMGNTTYTNAAYCAY 477
Qy 174 YGCNNAARAAAYACNCGNCAYTGGATNCARYTNGARAARATNMGNCNCCNWSNAYGT 233
Db 478 TGYTNGARCAAYTNGTNCARTAYTMGNACNGAYTGGAYCAYWSNTGACNGARCAWSN 537
Qy 234 NWSNACNGTNCATNCAYATNTNTAYTNCNCGARGCGCNAARGSGNNGTNCARTT 293
Db 538 GTNCAYTAYMGNCAAYARTTYSNTNCCNWSNNTNGAYGNCARARMGNTAYACNTTY 597
Qy 294 YCARTGGAARCAARGAYWSNTNMGNTNGGNGARGTNTAYGARGCNTGYTGG 345
Db 598 MGNCTNMGNWSMGNTTYAAYCCNTNTGYGGSNCGNCARCAITGGWSNG 649

RESULT 9
US-09-922-488-3
: Sequence 3, Application US/09922488
```

```
: Patent No. US20020119553A1
: GENERAL INFORMATION:
: APPLICANT: Thayer, Edward C.
: APPLICANT: Webster, Phillipa J.
: TITLE OF INVENTION: Human Secreted Protein, Zzpl
: FILE REFERENCE: 00-54
: CURRENT APPLICATION NUMBER: US/09/922,488
: CURRENT FILING DATE: 2001-08-03
: PRIOR APPLICATION NUMBER: 60/222,814
: PRIOR FILING DATE: 2000-08-04
: PRIOR APPLICATION NUMBER: 60/260,512
: PRIOR FILING DATE: 2001-01-09
: NUMBER OF SEQ ID NOS: 5
: SOFTWARE: FastSEQ for Windows Version 4.0
: SEQ ID NO 3
: LENGTH: 1905
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: degenerate nucleotide sequence
: NAME/KEY: misc_feature
: LOCATION: 6, 9, 12, 15, 18, 21, 24, 30, 36, 39, 42, 45, 48, 51, 54,
: LOCATION: 57, 60, 63, 66, 69, 72, 75, 78, 84, 90, 96, 99, 102, 105,
: LOCATION: 111, 123, 132, 141, 144, 147, 153, 156, 159, 162, 168, 171,
: LOCATION: 174, 183, 186, 198, 204, 213, 225, 243, 246, 249, 252
: OTHER INFORMATION: n = A,T,C or G
: NAME/KEY: misc_feature
: LOCATION: 255, 264, 267, 270, 276, 279, 288, 291, 300, 303, 315, 318,
: LOCATION: 327, 330, 333, 345, 348, 351, 354, 360, 363, 366, 372, 375,
: LOCATION: 384, 387, 390, 399, 405, 411, 414, 417, 420, 423, 429, 435,
: LOCATION: 438, 441, 444, 447, 456, 459, 462, 465, 468, 474, 477
: OTHER INFORMATION: n = A,T,C or G
: NAME/KEY: misc_feature
: LOCATION: 480, 486, 489, 492, 495, 498, 504, 507, 513, 516, 519, 525,
: LOCATION: 531, 534, 537, 540, 546, 549, 555, 558, 561, 567, 570, 573,
: LOCATION: 576, 579, 582, 585, 588, 591, 594, 597, 600, 603, 606, 609,
: LOCATION: 612, 615, 618, 624, 633, 636, 639, 654, 663, 675, 678
: OTHER INFORMATION: n = A,T,C or G
: NAME/KEY: misc_feature
: LOCATION: 684, 687, 705, 708, 711, 714, 720, 723, 732, 735, 738, 741,
: LOCATION: 744, 753, 765, 768, 786, 795, 798, 810, 816, 819, 822,
: LOCATION: 825, 837, 843, 852, 855, 858, 861, 864, 876, 879, 882, 888,
: LOCATION: 894, 897, 900, 912, 915, 921, 924, 927, 930, 936, 939
: OTHER INFORMATION: n = A,T,C or G
: NAME/KEY: misc_feature
: LOCATION: 942, 951, 957, 963, 966, 978, 981, 984, 993, 996, 999, 1008,
: LOCATION: 1011, 1014, 1023, 1041, 1044, 1047, 1050, 1068, 1071, 1077,
: LOCATION: 1080, 1086, 1089, 1095, 1098, 1107, 1113, 1116, 1122, 1131,
: LOCATION: 1134
: OTHER INFORMATION: n = A,T,C or G
: NAME/KEY: misc_feature
: LOCATION: 1143, 1146, 1155, 1158, 1167, 1170, 1173, 1176, 1179, 1182,
: LOCATION: 1185, 1191, 1197, 1200, 1203, 1206, 1209, 1212, 1218, 1221,
: LOCATION: 1227, 1239, 1245, 1248, 1257, 1272, 1278, 1281, 1284, 1287,
: LOCATION: 1290, 1296, 1299, 1305, 1311, 1314, 1317, 1320, 1326
: OTHER INFORMATION: n = A,T,C or G
: NAME/KEY: misc_feature
: LOCATION: 1329, 1335, 1341, 1344, 1347, 1350, 1353, 1368, 1371, 1374,
: LOCATION: 1377, 1380, 1386, 1398, 1407, 1413, 1416, 1422, 1428, 1437,
: LOCATION: 1443, 1449, 1452, 1461, 1464, 1467, 1473, 1476, 1479, 1482,
: LOCATION: 1491, 1503, 1509, 1512, 1515, 1518, 1524, 1527, 1530
: OTHER INFORMATION: n = A,T,C or G
: NAME/KEY: misc_feature
: LOCATION: 1536, 1539, 1542, 1548, 1551, 1554, 1557, 1560, 1563, 1566,
: LOCATION: 1572, 1581, 1584, 1587, 1590, 1599, 1602, 1605, 1608, 1614,
: LOCATION: 1620, 1623, 1626, 1632, 1635, 1638, 1641, 1644, 1647, 1653,
: LOCATION: 1656, 1659, 1662, 1665, 1671, 1680, 1683, 1686, 1689
: OTHER INFORMATION: n = A,T,C or G
: NAME/KEY: misc_feature
: LOCATION: 1701, 1704, 1707, 1710, 1713, 1716, 1719, 1722, 1734, 1740,
: LOCATION: 1749, 1752, 1755, 1758, 1761, 1764, 1770, 1776, 1782, 1785,
: LOCATION: 1788, 1791, 1794, 1797, 1800, 1806, 1809, 1812, 1815, 1818,
: LOCATION: 1821, 1824, 1827, 1830, 1833, 1836, 1839, 1842, 1848
```

```
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc_feature
; LOCATION: 1851, 1857, 1860, 1863, 1866, 1872, 1878, 1887, 1896, 1902
; OTHER INFORMATION: n = A,T,C or G
US-09-922-488-3

Query Match          9.9%; Score 34.8; DB 10; Length 1905;
Best Local Similarity 12.1%; Pred. No. 2.2;
Matches 34; Conservative 68; Mismatches 180; Indels 0; Gaps 0;

QY 5 ARTGYGNCACNATWATCAYCGNNAAYCGNCTNACNTTGTGAGCCNTAYGGCCNNGNG 64
Db 1275 HGTNMGNTNTNMGNGARCCNGTNCAYGTGARGTNGNTNTNCTNARMGNACNGAYCC 1334
QY 65 ARYTNACNACNACNTGYTTAAAYACNACNACNGCWNSNGNTNTNCTTYSNATNGNW 124
Db 1335 NAAYYTNGTNTNTNTNCAYCATGYTGGGNGCNCNWSNCCNNAAYCCNTTYCARCA 1394
QY 125 SNGGNWSNTGYMGNTTYSNTAYWSNGAYCCNWSNATNACNGTNTWSNTAYGCNNAARAAYA 184
Db 1395 RCCNCARTGGCCNATHVTNNSNGAYGGTGYCCNTTYAARGGNGAYWSNTAYMGNACNCA 1454
QY 185 AYACNGCNGAYTGATNCARTYNTCARARATNMGNCNCCNWSNAAAYGTNWSNACNGTNA 244
Db 1455 RATGGTNGCNTYNGAYGGNGCNCNCCNTTYCARWSNCAATAYCARMGNTTYACNGTNGC 1514
QY 245 TNCAYATNTNTAYTYTNCNCARGCARGCNAARGGNGARWSNG 286
Db 1515 NACNTTYGCNTNTNTNGAYWSNGNWSNCAARMGNGCNTYNG 1556

RESULT 10
US-09-864-761-7539
; Sequence 7539, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
```

```
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 7539
; LENGTH: 593
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC018583.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 7.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.7
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.9
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.9
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5.8
US-09-864-761-7539

Query Match          9.9%; Score 34.6; DB 10; Length 593;
Best Local Similarity 29.3%; Pred. No. 0.96;
Matches 58; Conservative 29; Mismatches 111; Indels 0; Gaps 0;

QY 41 TTTGYGARCCTAYGGCCNMGNGARYTNACNACNACNTGYTTNAAAYACNACNACNGCW 100
Db 89 TCATTTAGACTTCTAGCTCTATTACTTAAGAAATATATTTCTTAAGACTATAGCTGCCA 148
QY 101 SNGTNTNCARTTYSNATNGNWSNCGNWSNTGYMGNTTYSNTAYWSNGAYCCNWSNA 160
Db 149 TAGATAGTGATTTCTCTCAAGAAATCTGGCAAAAGTAAGTTGAAACCTTCTCTGGGGAAGAT 208
QY 161 TNACNGTNTWSNTAYGCNNAARAAYACNCGNCNGAYTGATNCARTYTNARAARATNMGNG 220
Db 209 TCACCATCTTAGATGACAGTAAGAAAATTTGTGATTCATTGGAAGAGGCTCAAGATACTA 268
QY 221 CNCCNWSNAAAYGTNWSNA 238
Db 269 CACTAACACAGAGTTTGA 286

RESULT 11
US-09-917-800A-1003
; Sequence 1003, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
```



; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (885)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (895)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (911)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-09-925-300-223

Query Match 9.6%; Score 33.6; DB 10; Length 921;  
Best Local Similarity 28.3%; Pred. No. 2.8;  
Matches 60; Conservative 34; Mismatches 118; Indels 0; Gaps 0;  
Qy 139 TTYNSNTAYNSGAYCCNWSNATNACNGTNSNTAYGNCNARAAAYAYACNGCNGAYTGG 198  
Db 652 TTCAGTGATTTAGATCAGAGAAAGCATTGGAAGTGATTCCTCAAGGTAGAGCAACAGCTGCT 711  
Qy 199 ATNCARYTNGARAARATNMGNCNCCNWSNAAAYGTNNNSNACNGTNNATNCAYATNTAY 258  
Db 712 AACAAACACGTCAGCTAGTAGAAACCGAAAGCCCTCAACTTTTCCCTATGCAGATT 771  
Qy 259 YTNCCNGARGAGCNAARGNGRWSNGTNCARTTYCARTGGAARCAARGAYWSNTNMGN 318  
Db 772 AATACTAACAGGAGCAAGAGTGCAATTTTACAAGTCCCCCAACAGAGGAAACGGTTGGG 831  
Qy 319 GTNGNGARGTNTAYGARGCNTGYTGCGCNTY 350  
Db 832 TTCAGCACAGTGTTAAAGGNTGTGTTGCTTT 863

## RESULT 15

US-09-815-242-6280  
; Sequence 6280, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011a  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6280  
; LENGTH: 1278  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: CDS

; LOCATION: (1)...(1278)  
US-09-815-242-6280

Query Match 9.5%; Score 33.4; DB 10; Length 1278;  
Best Local Similarity 25.7%; Pred. No. 4.2;  
Matches 58; Conservative 38; Mismatches 130; Indels 0; Gaps 0;  
Qy 31\* GCNGTNACNTTYTGYGARCCNTAYGGNCCNMGNGARVTNACNACNACNTGYTTNAAVACN 90  
Db 707 GCCGTACCGCTCGTGAAGCGCGCTGGGACAGCTTCAGTAATTTTGGCCCTGAATACG 766  
Qy 91 ACNACNCCNWSNGTNTNCARTTYWSNATNGNWSNGNWSNGTNGTNGTNGTNGTNGTNGT 150  
Db 767 ATCCCTCCGATCGTGTCTATCTTCTCTACCTGTGGCGTGTCTGTCTGTCTGTCTGTCT 826  
Qy 151 GAYCCNWSNATNACNGTNSNTAYGCNNAARAAAYAYACNCGNGAYTGGATNCARYTNGAR 210  
Db 827 TGTTTGTCAATTAACCGCTGCTGCGGATGCGGTGGCGCTGCGGTGGGAGCGTTGCGTG 886  
Qy 211 AARATNMGNCNCCNWSNAAAYGTNNNSNACNGTNNATNCAYATNTNT 256  
Db 887 AAGATGAAATCGCTGCGGTTCGCTGGGCTTAAGCCCGCTCGTAT 932

Search completed: November 7, 2002, 03:29:52  
Job time : 94 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 00:46:05 ; Search time 2216 Seconds  
(without alignments)  
2565.260 Million cell updates/sec

Title: US09897438BK-2

Perfect score: 351

Sequence: 1 GARCARTGYGGNACNATNAT.....AYGARGCNTGYTGGGCGNYTN 351

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estcom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	179	51.0	315	12	BE938667
c 2	96	27.4	600	12	BG803882
c 3	66.4	18.9	635	10	BB248113
c 4	55	15.7	568	12	BF387628
c 5	46.6	13.3	803	17	CNS018FX
c 6	44.4	12.6	935	17	CNS002X1
					AL098239 Drosophil

7	42.8	12.2	707	17	CNS04ALQ	AL282023 Tetraodon
8	42.8	12.2	1101	17	CNS00KK2	AL077673 Drosophil
c 9	42.6	12.1	855	17	CNS011PE	AL100556 Drosophil
10	42.2	12.0	1101	17	CNS0106X	AL098595 Drosophil
c 11	41.6	11.9	894	17	BH705629	BH705629 BOHWF68TR
12	41.6	11.9	1101	17	CNS00GDR	AL072060 Drosophil
13	41.4	11.8	978	17	CNS02JBA	AL199999 Tetraodon
c 14	41	11.7	1101	17	CNS00KK2	AL077673 Drosophil
c 15	40.4	11.5	746	12	BF694109	BF694109 602082720
c 16	40.4	11.5	844	13	BM400822	BM400822 5009-0-79
17	40.4	11.5	1101	17	CNS0037Q	AL064465 Drosophil
c 18	40.4	11.5	1101	17	CNS014T4	AL104578 Drosophil
19	40.2	11.5	803	17	CNS018FX	AL109287 Drosophil
20	40.2	11.5	1101	17	CNS00FOO	AL070854 Drosophil
21	40	11.4	972	17	CNS02BJJ	AL189928 Tetraodon
c 22	40	11.4	1101	17	CNS0142D	AL104803 Drosophil
23	39.8	11.3	555	17	AZ937550	AZ937550 2M0195H13
24	39.8	11.3	556	17	AZ937529	AZ937529 2M0195D15
c 25	39.8	11.3	732	17	CNS03NYB	AL252668 Tetraodon
c 26	39.8	11.3	1101	17	CNS00CNT	AL059413 Drosophil
27	39.6	11.3	894	12	BG441805	BG441805 GA_Ea001
28	39.4	11.2	650	12	BG404162	BG404162 602420202
c 29	39.4	11.2	1101	17	CNS0106D	AL098575 Drosophil
30	39.2	11.2	980	17	CNS00JMO	AL076572 Tetraodon
c 31	39.2	11.2	987	17	CNS035IN	AL071398 Drosophil
32	39.2	11.2	1100	17	CNS00G3S	AL228776 Tetraodon
c 33	39	11.1	306	17	AZ579300	AZ579300 1M0363N07
34	39	11.1	787	17	CNS010B7	AL098749 Drosophil
c 35	39	11.1	793	17	CNS012YX	AL102195 Drosophil
36	39	11.1	1101	17	CNS00L4E	AL067752 Drosophil
37	39	11.1	1101	17	CNS00KT0	AL077995 Drosophil
c 38	38.8	11.1	975	17	CNS00L8Q	AL068009 Drosophil
39	38.8	11.1	1033	17	CNS013FB	AL102785 Drosophil
c 40	38.8	11.1	1101	17	CNS00B2I	AL058538 Drosophil
c 41	38.8	11.1	1101	17	CNS00FOO	AL070854 Drosophil
42	38.6	11.0	624	14	BQ392162	BQ392162 NISC_mq22
c 43	38.6	11.0	787	17	CNS010B7	AL098749 Drosophil
44	38.6	11.0	1039	17	CNS006H7	AL064351 Drosophil
c 45	38.6	11.0	1204	17	CNS016E2	AL106628 Drosophil

#### ALIGNMENTS

RESULT 1 BE938667/c 315 bp mRNA linear EST 02-OCT-2000  
LOCUS QV0-TN0084-180800-342-a08 TN0084 Homo sapiens cDNA, mRNA sequence.  
DEFINITION BE938667  
ACCESSION BE938667.1 GI:10466224  
VERSION EST.  
KEYWORDS human.  
SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 315)  
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-QV0-TN0084-180  
800-342-a08&t3=2000-08-18&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 315.  
Location/Qualifiers  
1. .315

## FEATURES

source  
Location/Qualifiers  
1. .315  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="TN0084"  
/dev\_stage="Adult"  
/note="Organ: testis\_normal; Vector: puc18; Site\_1: Smar;  
Site\_2: Smal; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
BASE COUNT 84 a 64 c 78 g 89 t  
ORIGIN

Query Match 51.0%; Score 179; DB 12; Length 315;  
Best Local Similarity 53.8%; Pred. No. 5.6e-41;  
Matches 149; Conservative 61; Mismatches 67; Indels 0; Gaps 0;  
QY 74 CNACTGYTTNAAACACACAGCTTCTGTCFCCCAATTTCCATTTGGGTCAGGTTTCAT 133  
DB 312 CTAGGCCCTTATACAAACAGCTTCTGTCFCCCAATTTCCATTTGGGTCAGGTTTCAT 253  
QY 134 GYMNTTYSNTAYNSGAYCCNWSNATNACNGTNSNTAYGNCNAAAYAAACNGCNG 193  
DB 252 GTCCGCTTAGTTATTCAGACCCAGCATCATCGTGTATATATGCAAGATAACTCTCGG 193  
QY 194 AYTGGATNCARYTNGAARAATNMGNCNWSNAAAYGNTNWSNACNGTNCAYATNY 253  
DB 192 ACTGGATTACAGTACAGAAAATTAGACCCCTTCCCAATGTGACGCAATCATCCATATCC 133  
QY 254 TNTAYTNCNGARGARCCNAAAGGNGARWSNCTNCARTTYCARTTGAARCAARGAYWSNY 313  
DB 132 TCTACCTCTCGAGAGCCCAAGGGGAGAAATGTCCCAATTCAGTGGAGAGCAAGATC 73  
QY 314 TNMGNTNGNGARGTNTAYGARGCNTGTGGGCGNYT 350  
DB 72 TTCGTGTAGGTGAAGTGTATGAAGCCTGCTGGGCGCTT 36

RESULT 2  
BG803882/c  
LOCUS  
DEFINITION  
0243-03 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,  
mRNA sequence.

ACCESSION  
BG803882  
VERSION  
BG803882.1 GI:17950794  
KEYWORDS  
EST.  
SOURCE  
house mouse.  
ORGANISM  
Mus musculus

REFERENCE  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 600)

AUTHORS  
Mu, X., Zhao, S., Perishad, R., Hsieh, T.-F., Scarpa, A., Wang, S.W.,  
White, R.A., Beremand, P.D., Thomas, T.L., Gan, L. and Klein, W.H.  
TITLE  
Gene expression in the developing mouse retina by EST sequencing  
and microarray analysis

JOURNAL  
Nucleic Acids Res. 29 (24), 4983-4993 (2001)  
MEDLINE  
21671825  
COMMENT  
Contact: Klein WH  
Department of Biochemistry and Molecular Biology  
University of Texas M.D. Anderson Cancer Center  
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA  
Tel: 713 792 3546  
Fax: 713 790 0329.

FEATURES  
source

Location/Qualifiers  
1. .600  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone\_lib="Mouse E14.5 retina lambda ZAP II Library"  
/tissue\_type="neural retina"  
/dev\_stage="embryonic day 14.5 post-fertilization"  
/note="Vector: pAMP10 (Gibco); Cloned unidirectionally.  
Primer: Oligo dT. RNA Isolation: cytoplasmic RNA preps  
(Manniat); Cloning Technique: CUA Cloning (CloneAmp,  
Life Technologies); Average insert size: 1.8 Kb;  
Insertion site: TAGCTCCACTGAATCTGAGTG-->. Other  
information regarding entire library may be found at  
http://pga.swmed.edu/Data/Libraries/microarray\_cdna\_librar  
ies.htm."

BASE COUNT 127 a 167 c 152 g 152 t 2 others  
ORIGIN

Query Match 27.4%; Score 96; DB 12; Length 600;  
Best Local Similarity 58.6%; Pred. No. 1.1e-16;  
Matches 78; Conservative 30; Mismatches 25; Indels 0; Gaps 0;

QY 218 GNGCNCNWSNAAAYGTNWSNACNGTNCAYATNYTNTAYTNCNGARGCNCNARG 277  
DB 476 GAGCCCTTCCAATGTGAGCAGCATCCATCCATCTGCTACCTCCCGAGGAGCCAAAG 417  
QY 278 GNGARWSNCTNCARTTYCARTTGAARCAARGAYWSNCTNMGNTNGNGARGTNTAYGARG 337  
DB 416 GGGAGCGGTGCGAGTTCAGTGGNAAACAGACAGCCCTGCGAGTGGGTGTATGAGG 357  
QY 338 CNTGYTGGGCGNYT 350  
DB 356 CCGTGTGGGCGCT 344

RESULT 3  
BB248113  
LOCUS  
DEFINITION  
635 bp mRNA linear EST 23-OCT-2001  
musculus cDNA clone A730023J04 3', mRNA sequence.

ACCESSION  
BB248113  
VERSION  
BB248113.2 GI:16355611  
KEYWORDS  
EST.  
SOURCE  
house mouse.  
ORGANISM  
Mus musculus

REFERENCE  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 635)

AUTHORS  
Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A.,  
Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda,  
M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M.,  
Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki,  
D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H.,  
Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toyota, T.,  
Muramatsu, M. and Hayashizaki, Y.

TITLE  
RIKEN Mouse ESTs (Arakawa, T., et al. 2001)  
JOURNAL  
Unpublished (2001)

COMMENT  
On Jul 6, 2000 this sequence version replaced gi:8940859.  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsr.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh,  
M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,



Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)

Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences *Mamm. Genome* 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

**FEATURES**  
**SOURCE**

```

mouse tissues:
Location/Qualifiers
1. .635
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A730023J04"
/clone_lib="RIKEN full-length enriched, 7 days neonate
cerebellum"
/tissue_type="cerebellum"
/dev_stage="7 days neonate"
/lab_host="DH10B"
/note="Site_1: Sali; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA
primed with a primer [5'
GAGAGAGAGAGATCCAGCAAGCTCTTTTTTTTTTNN 3'], cDNA
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length
cap-trapper. cDNA went through one round of normalizat
to Rot = 20.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGAGATTCGAGTAAATTAATATCCGCCCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lam
PIC 1"

```

	TTC		
	a	c	g
BASE COUNT	182	122	118
ORIGIN			
			213

Query Match	18.9%	Score 66.4;	DB 10;	Length 635;
Best Local Similarity	57.3%;	Pred. No. 4.5e-08;		
Matches 55: Conservative	19;	Mismatches 22;	Indels 0;	Gaps 0;

QY 29 AYGCGTGNACNTTGTGYGARCCNTAYGGNCCNMGNGARYTNACNACNACNTGYTYTNAAYA 88

Db 1 ATGCTGTCACTTCTGTGAGCCGTACGGCCCTCGAGACCTGACCACCACATGCCTGAACA 60

QY 89 CNACNCGCNSGNTYTNCA<sup>RT</sup>TYWSNATNGGNW 124

Db 61 CAACAACAGCATCTGTCTCCAGTTTCCATTGGTA

RESULT 4  
BF387628/C

LOCUS	BF387628	568 bp	mRNA	linear	EST 27-NOV-2000
DEFINITION	UI-R-CAL-bbs-e-06-0-UI.s1 UI-R-CAL Rattus norvegicus cDNA clone				
ACCESSION	UI-R-CAL-bbs-e-06-0-UI 3', mRNA sequence.				
VERSION	BF387628	1	GI:11372452		

KEYWORDS  
SOURCE  
ORGANISM

EST.  
Norway rat.

*Rattus norvegicus*

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE  
AUTHORS

**TITLE** Normalization and subtraction: two approaches to facilitate gene discovery

**JOURNAL** Genome Res. 6 (9), 791-806 (1996)

**MEDLINE** 97044477

COMMENT

Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 6250  
Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized pons library cDNA library Preparation: M.B. Soares Lab

Clone distribution: clones will be available through Research Genetics ([www.resgen.com](http://www.resgen.com))

Seq primer: M13 Forward

[illegible]

FOUR-IES.  
Location/Qualifiers  
1. .568

1. .568

```

/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UT-R-CAL-bbs-e-06-0-UT"
/clone.lib="UT-R-CAL"
/lab_hosts="DH10B (Life Technolo
/notes="Vector: pWT3D-Pac (Phar
polylinker; Site_1: Not i; Site
library is a subtracted library
tissues: thalamus, cerebellum,
, midbrain, cerebral cortex, c
hippocampus. For a detailed de
from which this clone was deriv
site at ratseq.eng.uiowa.edu
previously described in (Bonaldi
Genome Research 6:791-806, 1996)
TAG_LTB-UT-R-CAL
TAG_TISSUE=pons
TAG_SEQ=AGCAGC"

```

BASE COUNT	116 a	156 c	131 q	165 t
rno_524-acbacc				

Query Match 15.7%; Score 55; DB 12; Length 568;  
Best Local Similarity 58.2%; pred. NO. 8.4e-05;  
Matches 46; Conservative 18; Mismatches 15; Indels

QY 272 CNAARGNGARWSNGTNCARTTYCARTGGAAARCARGAYWSNYTNMGNGTNGNGARGTNT 331

Db 567 CAAAGGGGAGGAACGTCCAGTTCAGTTCGAAACAGGACACCCCTGGCTGGGGTGGGTGT 508

QY 332 AYGARGCNTGYTGGGCNYT 350

Db 507 ACGAAGCCTGCTGGGCCCT 489

RESULT 5  
CNS018EY/C

CNSUI8FX/8	LOCUS	DEFINITION	ACCESSION
------------	-------	------------	-----------

CNS018FX 803 bp DNA linear GSS 26-JUL-1999  
Drosophila melanogaster genome survey sequence T7 end of BAC  
BACN13P04 of DrosBAC library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
AT109287

VERSION AL109287.1 GI:5629591  
KEYWORDS GSS.  
SOURCE Drosophila melanogaster.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 803)  
AUTHORS Genoscope.  
TITLE Direct Submission  
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)  
COMMENT - Web : www.genoscope.cns.fr  
Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.  
FEATURES  
source Location/Qualifiers  
1..803  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone="BACN13P04"  
/clone\_lib="DrosBAC"  
/plasmid="pBelobAC11"  
/note="end : T7"  
BASE COUNT 119 a 113 c 128 g 138 t 305 others  
ORIGIN  
Query Match 13.3%; Score 46.6; DB 17; Length 803;  
Best Local Similarity 12.2%; Pred. No. 0.03;  
Matches 30; Conservative 102; Mismatches 114; Indels 0; Gaps 0;  
Qy 92 CNACGWNSTNTNCARTYTSNATNGNWSNGNWSNTGYMNTYTSNTAYWSNG 151  
Db 756 CBBSSNNKSKKKCCBKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 697  
Qy 152 AYCCNSNATNACNTNSNTAYGCNAAAYAAACNGCNGAYTGGATNCARYTGARA 211  
Db 696 KCCCCCBKK 637  
Qy 212 ARATNMGNCNCCNSNAAYGTNWSNACNGTNCATNCAYATNTYNTAYTNCNGARGARG 271  
Db 636 KKKMHMKCCCKMTADAKKKKKMAKKDKADACCCMKAAKKAAKKKKKKKKKA 577  
Qy 272 CNAARGNGARWSNCTNCARTYTCARTGGAARCARGATWSNTYNNMGNTNGNGARGTNT 331  
Db 576 AKTKAKKKKKNNNNKMAAKAAKAKKACDCAADAAADAAVKKAAAKKAAKMAKSGRAK 517  
Qy 332 AYGARG 337  
Db 516 KKGKKK 511  
RESULT 6  
CNS002X1 935 bp DNA linear GSS 26-JUL-1999  
LOCUS Drosophila melanogaster genome survey sequence SP6 end of BAC  
DEFINITION BACN02A14 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
ACCESSION AL098239  
VERSION AL098239.1 GI:5609850  
KEYWORDS GSS.  
SOURCE Drosophila melanogaster.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 935)  
AUTHORS Genoscope.

TITLE Direct Submission  
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)  
COMMENT - Web : www.genoscope.cns.fr  
Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.  
FEATURES  
source Location/Qualifiers  
1..935  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone="BACN02A14"  
/clone\_lib="DrosBAC"  
/plasmid="pBelobAC11"  
/note="end : SP6"  
BASE COUNT 95 a 36 c 129 g 307 t 368 others  
ORIGIN  
Query Match 12.6%; Score 44.4; DB 17; Length 935;  
Best Local Similarity 11.9%; Pred. No. 0.15;  
Matches 38; Conservative 123; Mismatches 158; Indels 0; Gaps 0;  
Qy 20 TCCAYGGNAAYCGCTNACNTTYTGYGARCCTNATYGGNCCNNGARCTNACNACNACNT 79  
Db 617 TTMMGTMSVTAHSGMMMTGCTTMMKMGMTTTCVTGTMGTATTMTTAMMM 676  
Qy 80 GYTTNAAACNACNACNCGNNSNGTNTNCARTTYWSNATNGNWSNGNWSNTGYMNT 139  
Db 677 STGGCMMTMTACMNMNMGTRMMHNVTVTVTKRSAGSACGKSGGGTGWGGT 736  
Qy 140 TYWSNTAYWSNGAYCCNWSNATNACNGTNSNTAYGCNAAAYAAAYACNGCNGAYTGA 199  
Db 737 TKACTVAMTVMTVBTATGGTGCAMCTGMMMTACGMGTVMMTTGMGTGAVGAGTKT 796  
Qy 200 TNCARYTNGARAARATNMGNCNCCNWSNAYAGTNNWSNACNCTNATNCAYATNTYNTAY 259  
Db 797 KTTATTTTGTAMGAGRRMTVGVDTTGHVAVTMMRRGSKTGGAARTNRNGGAATAMT 856  
Qy 260 TNCNGARGACNAAARGNGARWSNGTNCARTTYTCARTGGAARCARGATWSNTYNNMG 319  
Db 857 KRMTHHTKTGTTTTHSACAGVNVSSGMMVHBTGWMGTGHTVAGTGYMMMMHMRVMT 916  
Qy 320 TNGNGARGTNTAYGARGC 338  
Db 917 KGRMTVAMCMSAMGMRM 935  
RESULT 7  
CNS04ALQ 707 bp DNA linear GSS 21-MAY-2000  
LOCUS Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone  
DEFINITION 095B14 of library G from Tetraodon nigroviridis, genomic survey sequence.  
ACCESSION AL282023  
VERSION AL282023.1 GI:8020357  
KEYWORDS GSS; genome survey sequence.  
SOURCE Tetraodon nigroviridis.  
ORGANISM Tetraodon nigroviridis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.  
REFERENCE 1 (bases 1 to 707)  
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.  
TITLE Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence



d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.

FEATURES		Location/Qualifiers	
source		1..855	
		/organism="Drosophila melanogaster"	
		/db_xref="taxon:7227"	
		/clone="BACN06M20"	
		/clone_lib="DrosBAC"	
		/plasmid="pBelOBAC11"	
		/note="end : sp6"	
BASE COUNT	96 a	63 c	75 g
ORIGIN		160 t	461 others
Query Match	12.1%	Score 42.6;	DB 17; Length 855;
Best Local Similarity	18.5%,	Pred. No. 0.46;	Indels 4; Gaps 1;
Matches	60;	Conservative 98;	Mismatches 163;
Qy	22	CAYGSAAYGCNGTGNACNTTTGTGGARCGTAATYAGNCNCNMGMNGARTYNACNACNACNTGY	81
Db	771	CRCKMKKKCCSMCCCKCBCCGSCBCCKKCCSSKCMCKCCKCGAKKMACMCBC	712
Qy	82	YTNAAYACNACNACGCMWSNGTNYTCNARTYTSNATNGGNSNGGWSNMGNTTY	141
Db	711	KCKWDMCKMKCKCKCKBKFKMAACMCKNNKNKKKKNNKNNKNNNNNNNNNNK	652
Qy	142	WSNTAYSNGAYCCNWSNATNACGTNWSNTAYGCAARAAYAAVACNCNGAYTGGA	201
Db	651	NNKKNN	592
Qy	202	CARYTNGAARAATNMGNCGCCNWSNAACTNWSNACNGTNNATNCAYATNTYTAYTN	261
Db	591	AAGK---KAAAATAAKNGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	536
Qy	262	CCNGARGACNAARGGNGARWSNGTNCARTYTCYTCGGAARCARGAYWSNWTNG	321
Db	535	KNKGANKKKKKNNKNKKNNKKNNKKKKNNNNNNNNNNNNNNNNNNNNNNNN	476
Qy	322	GGNGARTNTAYGARGCNTGYTGGG	346
Db	475	NNKKNN	551
RESULT 10			
LOCUS	CNS0106X	1101 bp	DNA linear
DEFINITION	BACN03K20 of DrosBAC library from Drosophila melanogaster (fruit fly); genomic survey sequence.	GSS 26-JUL-1999	T7 end of BAC
ACCESSION	AL098595		
VERSION	AL098595.1	GI:5610206	
KEYWORDS	GSS.		
SOURCE	Drosophila melanogaster.		
ORGANISM	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
REFERENCE	1 (bases 1 to 1101)		
AUTHORS	Genoscope.		
TITLE	Direct Submission		
JOURNAL	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage ; BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr		
COMMENT	- web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk/. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.		
FEATURES		Location/Qualifiers	

```

source
1. .1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACN03K20"
/clone_lib="DrosBAC"
/plasmid="pBelobAC11"
/note="end : T7"

BASE COUNT      258 a   107 c    60 g   175 t   501 others
ORIGIN

Query Match          12.0%; Score 42.2; DB 17; Length 1101;
Best Local Similarity 9.7%; Pred. NO. 0.73;
Matches 31; Conservative 148; Mismatches 141; Indels 1; Gaps 1;

Qy 1 GARCAATGCGGNACNATNATGCAYGGNAAYCNCGTNAGTNTTTTGTCGARCCNTAYGGNCCN 60
   |: :::: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 779 DADADAADAKADAKADADADDGRCDGCKKKRAKDEKKKKDKAWGDKKKAKKDKAA 838
   |: :::: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

Qy 61 MGNGARTYNACNACNACTGYTYTATYACNACNACNCGNSNGTNYTNCARTTTYWSNATN 120
   |: :::: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 839 AKAKADAAGDAKARRRAGDRDKADAKAADAACAATAKAKKDDDDDAKATKAKTKAKDKAK 898
   |: :::: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

Qy 121 CGNHSNGCNWNTGYMGNTTYSWTATYSGNAYCCNWSNATNACNGTNWSWTAYGCNNAAR 180
   |: :::: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 899 AKKKKKDKDKAKAKAKADKDDKDDKDDKDDKDDKDDKDDKDDKDDKDDKDDKDDKDDK 958
   |: :::: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

Qy 181 AYAAYACGNCNGAYTGATNCARYTNGARA-ARATNMGNCNCNWSNAYGTNWSNAC 239
   |: :::: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 959 DKAKDKDKDDAKAWAKAKADADAKAKAKAKAKAADDDAAAADAKAADAADAKAKAK 1018
   |: :::: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

Qy 240 NGTNATNCAYATNTTAYTYTNCNGARGCGCNAARGGNGSNATNCARTTYCARTG 299
   |: :::: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 1019 ADDAAKAKADAAADADAKAAKAAKAAKADADAADAKAKAKADDDKDDKDDKDDADKAK 1078
   |: :::: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

Qy 300 GAARCARGAYSWNTNMGNT 320
   |: :::: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 1079 KAKAKADAKAKAKAKAKAK 1099
   |: :::: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :


RESULT 11
BH705629/c
LOCUS       BH0WF68TR BO_2_3_KB Brassica oleracea genomic clone BOHWF68, DNA
DEFINITION sequence.
ACCESSION   BH705629
VERSION     BH705629.1 GI:18788501
KEYWORDS    GS.
SOURCE      Brassica oleracea.
ORGANISM    Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1 (bases 1 to 894)
AUTHORS    Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
TITLE       Whole genome shotgun sequencing of Brassica oleracea
JOURNAL     Unpublished (2001)
COMMENT     Other_GSSs: BOHWF68TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source           1..894
                     /organism="Brassica oleracea"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOHWF68"
                     /clone_lib="BO_2_3_KB"

```





Db 343 TCAATGGCCACATTAAATAACCAAAATAATTAACACACAGAAAGTTAAACCTACAATACATT 284  
Qy 167 TNWSNTAYGCNAARAAYAAAYACNGCNGAYTGGATNCARYTNGARAARATNMGNGCNCNW 226  
Db 283 TCACACATCTACTGTGTAAACCACCAAAATTTTAAAAAATTTTAAAAATATTTTATACCTT 224  
Qy 227 SNAAYGTNWSNACNGTNAATNCAYATNTYTNCCNGARGARGCNAARGGNGGARWSNG 286  
Db 223 TAAATATTAAAAAATATTTTAAAAATGGCTATGCAAGAAAAAAGGAAAAAGAAATACAT 164  
Qy 287 TNCARTTYCARTGGAARCARGAYWSNYTNMGNGTNGNGAR 327  
Db 163 TTCATTACCATATATAACACACACTTTAAAAATGGGTAGAA 123

Search completed: November 7, 2002, 02:17:01  
Job time : 2223 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 6, 2002, 22:05:00 ; Search time 2212 Seconds  
(without alignments)  
856.633 Million cell updates/sec

Title: US-09-897-438B-2

Perfect score: 117

Sequence: 1 EQCGTTHMGNAVTFCEPYGP.....FQWKQDSLVRGEVYACWAL 117

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Word size: 1

Total number of hits satisfying chosen parameters: 32308016

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-Q/cgna2\_1/USPTO.spool/US09897438/runat\_06112002\_101956\_11811/app\_query.fasta\_1.263  
-DB=EST -QFMT=fastap -SUFFIX=oligo.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09897438\_@CGN1\_1\_1716\_@runat\_06112002\_101956\_11811 -NCPU=6 -ICPU=3  
-NO\_XLPAY -NO\_MMAPP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREDS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPO=6 -FGAPEXT=7  
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_man:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
c 1	44	37.6	600	12	BG803882
c 2	31	26.5	635	10	BG803882 0243-03 M
c 3	27	23.1	315	12	BE248113
c 4	13	11.1	568	12	BE938667
c 5	9	7.7	717	17	BF387628 UI-R-CAL-
c 6	8	6.8	260	17	AQ688885 nxb0078D
c 7	8	6.8	308	10	AZ010548
c 8	8	6.8	312	12	AW114381 rs51810.Y
c 9	8	6.8	354	10	BF642099 NF039F071
c 10	8	6.8	366	17	AW911732
c 11	8	6.8	371	14	AZ791276 2M0041F09
c 12	8	6.8	381	17	C47012 C47012 Yuj1
c 13	8	6.8	381	17	AZ033696
c 14	8	6.8	399	17	AQ087955
c 15	8	6.8	400	10	AQ030475
c 16	8	6.8	400	10	AW696758
c 17	8	6.8	411	10	AW789434
c 18	8	6.8	411	10	AV803611
c 19	8	6.8	421	9	AW655178
c 20	8	6.8	451	10	AA758580
c 21	8	6.8	456	12	BB825159
c 22	8	6.8	457	13	BF775130
c 23	8	6.8	475	17	BM105816
c 24	8	6.8	476	9	AQ994137
c 25	8	6.8	478	12	AL369916
c 26	8	6.8	479	10	BF639389
c 27	8	6.8	489	17	AW668785
c 28	8	6.8	491	10	AQ603779
c 29	8	6.8	493	17	AW684418
c 30	8	6.8	501	9	AQ345388
c 31	8	6.8	503	17	AL367517
c 32	8	6.8	509	9	AQ669357
c 33	8	6.8	519	13	AU207398
c 34	8	6.8	524	12	BI589981
c 35	8	6.8	528	12	BF641085
c 36	8	6.8	534	17	BE08142
c 37	8	6.8	539	17	AQ562548
c 38	8	6.8	546	12	AQ773077
c 39	8	6.8	549	10	BE754467
c 40	8	6.8	549	13	AW669102
c 41	8	6.8	550	17	BI775248
c 42	8	6.8	551	17	AL765609
c 43	8	6.8	552	9	AQ298583
c 44	8	6.8	557	12	AI997627
c 45	8	6.8	566	12	BE756494
					BF005427 EST433925

# ALIGNMENTS

RESULT 1	BG803882/c	600 bp	linear	EST 20-DEC-2001
LOCUS	0243-03 Mouse E14.5 retina lambda ZAP II Library Mus musculus cdna,			
DEFINITION	mRNA sequence.			
ACCESSION	BG803882			
VERSION	BG803882.1			
KEYWORDS	EST			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
	1 (bases 1 to 600)			
	Mu, X., Zhao, S., Pershad, R., Hsieh, T.-F., Scarpa, A., Wang, S.W.,			
	White, R.A., Beremand, P.D., Thomas, T.L., Gan, L. and Klein, W.H.			

**TITLE** Gene expression in the developing mouse retina by EST sequencing and microarray analysis

**JOURNAL** Nucleic Acids Res. 29 (24), 4983-4993 (2001)

**MEDLINE** 21671825

**COMMENT** Contact: Klein WH  
Department of Biochemistry and Molecular Biology  
University of Texas M.D. Anderson Cancer Center  
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA  
Tel: 713 792 3646  
Fax: 713 790 0329.

**FEATURES** Location/Qualifiers

**Source** 1..600  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone\_lib="Mouse E14.5 retina lambda ZAP II Library"  
/tissue\_type="neural retina"  
/dev\_stage="embryonic day 14.5 post-fertilization"  
/note="Vector: pAMP10 (Gibco); Cloned unidirectionally.  
Primer: Oligo dt. RNA isolation: cytoplasmic RNA preps  
(Mannatis); Cloning Technique: CUA Cloning (CloneAmp,  
Life Technologies); Average insert size: 1.8 Kb;  
Insertion site: TAGTCCACTGAATCTGAGTG-->. Other  
information regarding entire library may be found at  
http://pga.swmed.edu/Data/Libraries/microarray\_cdna\_librar  
ies.htm."

**BASE COUNT** 127 a 167 c 152 g 152 t 2 others

**ORIGIN**

**Alignment Scores:**

**Pred. No.:** 6.58e-37 **Length:** 600

**Score:** 44.00 **Matches:** 44

**Percent Similarity:** 100.00% **Conservative:** 0

**Best Local Similarity:** 100.00% **Mismatches:** 0

**Query Match:** 37.61% **Indels:** 0

**DB:** 12 **Gaps:** 0

US-09-897-438B-2 (1-117) x BG803882 (1-600)

**QY** 74 AlaProSerAsnValSerThrValIleHisIleLeuTyrLeuProGluGluAlaLysGly 93  
|||||  
**Db** 474 GCCCTTCCATGTGAGCAGCATCTCCATCTGTACTTCCCGAGAGCCAAAGG 415  
|||||

**QY** 94 GluSerValGlnPheGlnTrpLysGlnAspSerLeuArgValGlyGluValTyrGluAla 113  
|||||  
**Db** 414 GAGAGCGTGCAGTTCACAGTGGAAACAGGACAGCGCTGCGAGTGGGTGAGGTGTATGAGGCC 355  
|||||

**QY** 114 CysTrpAlaLeu 117  
|||||

**Db** 354 TGCTGGGCCCTG 343

**RESULT** 2

**BB248113**

**LOCUS** BB248113

**DEFINITION** BB248113 RIKEN full-length enriched, 7 days neonate cerebellum Mus musculus cDNA clone A730023J04 3', mRNA sequence.

**ACCESSION** BB248113

**VERSION** BB248113.2 GI:16355611

**KEYWORDS** EST.

**SOURCE** house mouse.

**ORGANISM** Mus musculus

**REFERENCE** Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 635)

**AUTHORS** Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toyota,T., Muramatsu,M. and Hayashizaki,Y.

**TITLE** RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

**JOURNAL** Unpublished (2001)

**COMMENT** On Jul 6, 2000 this sequence version replaced gi:8940859.

Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp,  
URL:http://genome.gsc.riken.go.jp/  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.  
Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

**FEATURES** Location/Qualifiers

**Source** 1..635  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone\_lib="A730023J04"  
/clone\_lib="RIKEN full-length enriched, 7 days neonate cerebellum"  
/tissue\_type="cerebellum"  
/dev\_stage="7 days neonate"  
/lab\_host="DH10B"  
/note="Site\_1: SalI; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'  
GAGAGAGAGAGATCCAGAGCTTTTTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCGAGTTAATTAATTAATCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

**BASE COUNT** 182 a 122 c 118 g 213 t

**ORIGIN**

**Alignment Scores:**

**Pred. No.:** 6.55e-23 **Length:** 635

**Score:** 31.00 **Matches:** 31

**Percent Similarity:** 100.00% **Conservative:** 0

**Best Local Similarity:** 100.00% **Mismatches:** 0

Query Match:	26.50%	Indels:	0
DB:	10	Gaps:	0
US-09-897-438b-2 (1-117) x BB248113 (1-635)			
QY	11	AlaValThrPheCysGluProTyrGlyProArgGluLeuThrThrCysLeuAsnThr	30
Db	3	GCTGTCACTTCTGTGAGCGTACGCCCTCGAGAGCTGACCACCATGCCTGAACACA	62
QY	31	ThrThrAlaSerValLeuGlnPheSerIleGly	41
Db	63	ACAACAGCATGTCTCCAGTTTCCCATGGT	95
RESULT 3			
LOCUS	BE938667/c	315 bp	mRNA linear EST 02-OCT-2000
DEFINITION	QV0-TN0084-180800-342-a08 TN0084 Homo sapiens cDNA, mRNA sequence.		
ACCESSION	BE938667		
VERSION	BE938667.1		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bai,G.S., Simpson,D.H., Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.		
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)		
MEDLINE	20202663		
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-QV0-TN0084-180800-342-a08&t3=2000-08-18&t4=1) Seq primer: puc 18 forward High quality sequence stop: 315.		
FEATURES	Location/Qualifiers		
source	1..315		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone_lib="TN0084"		
	/dev_stage="Adult"		
	/note="Organ: testis_normal; Vector: puc18; Site.1: SmaI; Site.2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No.196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."		
BASE COUNT	84 a	64 c	78 g
ORIGIN	89 t		
Alignment Scores:			
Pred. No.:	5,21e-19	Length:	315
Score:	27.00	Matches:	27
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	23.08%	Indels:	0
DB:	12	Gaps:	0
US-09-897-438b-2 (1-117) x BE938667 (1-315)			
QY	28	LeuAsnThrThrThrAlaSerValLeuGlnPheSerIleGlySerGlySerCysArgPhe	47
Db	304	CTTAATACACACACAGCTTCTGCTCCCAATTTTCCATTGGGTGAGGTCATGCTCGCTT	245
QY	48	SerTyrSerAspProSerIle	54
Db	244	AGTTATTACAGACCCAGCATC	224
RESULT 4			
LOCUS	BF387628/c	568 bp	mRNA linear EST 27-NOV-2000
DEFINITION	UI-R-CAL-bbs-e-06-0-UI.s1 UI-R-CAL Rattus norvegicus cDNA clone		
ACCESSION	BF387628		
VERSION	BF387628.1		
KEYWORDS	EST.		
SOURCE	Norway rat.		
ORGANISM	Rattus norvegicus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
AUTHORS	1 (bases 1 to 568)		
TITLE	Normalization and subtraction: two approaches to facilitate gene discovery		
JOURNAL	Genome Res. 6 (9), 791-806 (1996)		
MEDLINE	97044477		
COMMENT	Contact: Soares, MB Program for Rat Gene Discovery and Mapping University of Iowa 451 Eckstein Medical Research Building Iowa City, IA 52242, USA Tel: 319 335 8250 Fax: 319 335 9565 Email: msoares@blue.weeg.uiowa.edu The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NciI site and the oligo-dT track served to identify it as a clone from the normalized pons library cDNA library preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com) Seq primer: M13 Forward POLYA=Yes.		
FEATURES	Location/Qualifiers		
source	1..568		
	/organism="Rattus norvegicus"		
	/strain="Sprague-Dawley"		
	/db_xref="taxon:10116"		
	/clone_lib="UI-R-CAL-bbs-e-06-0-UI"		
	/clone_lib="UI-R-CAL"		
	/lab_host="DH10B (Life Technologies)"		
	/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker: Site.1: Not I; Site.2: Eco RI; The UI-R-CAL library is a subtracted library derived from the following tissues: thalamus, cerebellum, hypothalamus, medulla, pons, midbrain, cerebral cortex, corpus striatum, testis, and hippocampus. For a detailed description of the library from which this clone was derived, please visit our web site at <a href="http://ratest.eng.uiowa.edu">ratest.eng.uiowa.edu</a> . The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)		
	TAG_LIB=UI-R-CAL		
	TAG_TISSUE=pons		
	TAG_SEQ=AGCAGC"		
BASE COUNT	116 a	156 c	131 g
ORIGIN	165 t		
Alignment Scores:			
Pred. No.:	0.00123	Length:	568
Score:	13.00	Matches:	13

Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 11.11% Indels: 0  
 DB: 12 Gaps: 0

US-09-897-438B-2 (1-117) x BF387628 (1-568)

QY 105 LeuArgValGlyuValTyGluAlaCysTriPAlaLeu 117  
 Db 526 CTCGCTGGGTGAGGTGTACGAGCCTGCTGGCCCTG 488

# RESULT 5

AQ688885 717 bp DNA linear GSS 01-JUL-1999  
 LOCUS nbxb0078D03f CUGI Rice BAC Library Oryza sativa genomic clone  
 DEFINITION nbxb0078D03f, DNA sequence.

ACCESSION AQ688885  
 VERSION GI:5330053  
 KEYWORDS GSS.

SOURCE Oryza sativa.  
 ORGANISM Oryza sativa

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.

AUTHORS Wing, R.A. and Dean, R.A.  
 TITLE A BAC End Sequencing Framework to Sequence the Rice Genome

JOURNAL Unpublished (1998)  
 COMMENT Contact: Wing RA

Clemson University Genomics Institute  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Tel: 864 656 7288  
 Fax: 864 656 4293

Email: twing@clemson.edu

Seq primer: TAATACGACTCACTATAGG

Class: BAC ends

High quality sequence stop: 305.

# FEATURES

SOURCE

1..717  
 Location/Qualifiers  
 /organism="Oryza sativa"  
 /strain="Japonica"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:4530"  
 /clone="nbxb0078D03f"  
 /clone\_lib="CUGI Rice BAC Library"  
 /tissue\_type="Leaf"  
 /lab\_host="E. coli DH10B"  
 /note="Vector: pBeloBAC11; Site\_1: HindIII; Site\_2: HindIII; Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 Kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."

175 a 160 c 164 g 218 t  
 BASE COUNT  
 ORIGIN

Alignment Scores:  
 Pred. No.: 33.1 Length: 717  
 Score: 9.00 Matches: 9

Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 7.69% Indels: 0  
 DB: 17 Gaps: 0

US-09-897-438B-2 (1-117) x AQ688885 (1-717)

QY 41 GlySerGlySerCysArgPheSerTyr 49  
 Db 359 GGAAGCGGTTCAATGCCGTTTCTCTAC 385

# RESULT 6

AZ010548 260 bp DNA linear GSS 25-FEB-2000  
 LOCUS RPCI-23-325B10.TJ RPCI-23 Mus musculus genomic clone RPCI-23-325B10  
 DEFINITION , DNA sequence.

ACCESSION AZ010548  
 VERSION AZ010548.1 GI:7085932  
 KEYWORDS GSS.

SOURCE house mouse.  
 ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 1 (bases 1 to 260)

AUTHORS Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret

, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.  
 and Fraser, C.M.

Mouse BAC End Sequences from Library RPCI-23

Unpublished (1999)

Other\_GSSs: RPCI-23-325B10.TV

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics ([info@resgen.com](http://info@resgen.com)). BAC end page: [http://www.tigr.org/tldb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html)

Plate: 325 row: B column: 10

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

source

1..260  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPCI-23-325B10"  
 /clone\_lib="RPCI-23"  
 /sex="Female"  
 /lab\_host="DH10B"  
 /note="Organ: Kidney/Brain; Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

72 a 46 c 63 g 79 t  
 BASE COUNT  
 ORIGIN

Alignment Scores:  
 Pred. No.: 105 Length: 260  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 6.84% Indels: 0  
 DB: 17 Gaps: 0

US-09-897-438B-2 (1-117) x AZ010548 (1-260)

QY 41 GlySerGlySerCysArgPheSer 48  
 Db 217 GGAAGTGGCTCCTGTAGATTACG 240

RESULT 7  
 LOCUS AW114381 308 bp mRNA linear EST 10-MAY-2001  
 DEFINITION rs51b10.y1 Sommer Pristionchus Pristionchus pacificus cDNA 5', mRNA sequence.

ACCESSION AW114381  
 VERSION AW114381.1 GI:6080719  
 KEYWORDS EST.  
 SOURCE Pristionchus pacificus.  
 ORGANISM Pristionchus pacificus.

REFERENCE 1 (bases 1 to 308)  
 AUTHORS Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Neodiplogasteridae; Pristionchus.

McCarte, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T., Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagarelis, W., R., Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe, M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.  
 TITLE The Washington Univ. Nematode EST Project, 1999  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: McCarter JP  
 The Washington Univ. Nematode EST Project, 1999  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu

The library was constructed by Dr. Ralf Sommer DNA Sequencing by:  
 Washington University Genome Sequencing Center  
 Contact Dr. Ralf Sommer (ralf.sommer@uebingen.mpg.de) for  
 information about this clone.  
 Putative full length read  
 The vector to vector length is 374  
 Seq primer: -40RP from Gibco  
 High quality sequence stop: 307.

FEATURES  
 Location/Qualifiers  
 1..308  
 /organism="Pristionchus pacificus"  
 /strain="PS 312"  
 /db\_xref="taxon:54126"  
 /clone\_lib="Sommer Pristionchus"  
 /sex="predominantly hermaphroditic"  
 /dev\_stage="mixed stages (embryo to adult)"  
 /lab\_host="not applicable (host cell line)"  
 /notes="Vector: Uni-ZAP XR vector (Stratagene); Site\_1: 5' EcoRI; Site\_2: 3' XhoI; 1st strand cDNA was primed with a XhoI - oligo(dT) primer. Double-stranded cDNA was ligated to EcoRI adaptors digested with XhoI and cloned into XhoI and EcoRI sites. Primary complexity of the library was 10 in the 7th. The library went through one round of amplification."

BASE COUNT 69 a 92 c 38 g 108 t 1 others  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 131 Length: 308  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 6.84% Indels: 0  
 DB: 10 Gaps: 0

US-09-897-438B-2 (1-117) x AW114381 (1-308)

QY 50 SerAspProSerIleThrValSer 57

Db 272 TCGACCCATCCATTACAGTATCC 295

RESULT 8  
 LOCUS BF642099/c 312 bp mRNA linear EST 19-DEC-2000  
 DEFINITION NF039F07IN1062 Insect herbivory Medicago truncatula cDNA clone sequence.

ACCESSION BF642099  
 VERSION BF642099.1 GI:11906257  
 KEYWORDS EST.  
 SOURCE barrel medic.  
 ORGANISM Medicago truncatula

REFERENCE 1 (bases 1 to 312)  
 AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.

Korth, K., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J., Flores, H.R., Inman, J.T., Weller, J.W. and May, G.D.  
 TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation  
 JOURNAL Medicago truncatula insect herbivory library  
 COMMENT Unpublished (2000)  
 Contact: Korth K  
 Dept. of Plant Pathology  
 University of Arkansas  
 217 Plant Science Building, Fayetteville, AR 72701, USA  
 Tel: 501 575 5191  
 Fax: 501 575 7601  
 Email: korth@comp.uark.edu  
 Insert Length: 312 Std Error: 0.00  
 Plate: 039 row: F column: 07  
 Seq primer: TCACACAGGAACAGCTATGAC.

FEATURES  
 Location/Qualifiers  
 1..312  
 /organism="Medicago truncatula"  
 /db\_xref="taxon:3880"  
 /clone\_lib="NF039F07IN"  
 /clone\_lib="Insect herbivory"  
 /tissue\_type="local and systemic leaves"  
 /note="Vector: Lambda Zap; Library was produced from fully expanded M. truncatula leaves of plants fed upon by Spodoptera exigua (beet armyworm) for 24 hours. Systemic (undamaged leaves from injured plants) and wounded leaves were harvested and pooled."

BASE COUNT 69 a 69 c 84 g 90 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 133 Length: 312  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 6.84% Indels: 0  
 DB: 12 Gaps: 0

US-09-897-438B-2 (1-117) x BF642099 (1-312)

QY 74 AlaProSerAsnValSerThrVal 81  
 Db 289 GCTCCCTCCATGTCACAGGTC 266

RESULT 9

LOCUS AW911732 354 bp mRNA linear EST 25-MAY-2000  
 DEFINITION uf38f01.v1 Soares\_mammary\_gland\_NMLMG Mus musculus cDNA clone sequence.

ACCESSION AW911732  
 VERSION AW911732.1 GI:8077314  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus

US-09-897-438B-2 (1-117) x AW114381 (1-308)

QY 50 SerAspProSerIleThrValSer 57

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 354)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Other\_ESTs: uf38f01.x1

Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
MGI:940509

Seq primer: -40RP from Gibco  
High quality sequence stop: 332.

FEATURES  
source

1. .354  
Location/Qualifiers  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1513657"  
/clone\_lib="Soares\_mammary\_gland\_NMLMG"  
/sex="female (lactating)"  
/tissue\_type="mammary gland"  
/lab\_host="DH10B"

/note="vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from mammary gland tissue from a lactating female, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Ronaldo."

BASE COUNT 68 a 99 c 118 g 69 t

Alignment Scores:

Pred. No.: 157 Length: 354  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 6.84% Indels: 0  
DB: 10 Gaps: 0

US-09-897-438B-2 (1-117) x AW911732 (1-354)

Qy 39 SerIleGlySerGlySerCysArg 46

Db 7 TCCATCGGCTCTGCGAGCTGCAGG 30

RESULT 10

AZ791276/c

LOCUS

DEFINITION AZ791276 366 bp DNA linear GSS 16-FEB-2001  
2M0041F09F Mouse 10kb plasmid UUGC1M library Mus musculus genomic

clone UUGC2M0041F09 F, DNA sequence.

ACCESSION AZ791276

VERSION AZ791276.1 GI:12933997

KEYWORDS

SOURCE

ORGANISM

house mouse.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 366)

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,

Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly

,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.

and Wright,D. Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: [ddunn@genetics.utah.edu](mailto:ddunn@genetics.utah.edu)

Insert Length: 10000 Std Error: 0.00

Plate: 0041 row: F column: 09

Seq primer: CGTTGTAAACGACGCGCAGT

Class: plasmid ends

High quality sequence stop: 366.

FEATURES

Location/Qualifiers

source

1. .366

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC2M0041F09"

/clone\_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(<http://www.jax.org/resources/documents/dnares/>). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of pWB42 (gil14732114(gb)AF129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to

adapted vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance."

BASE COUNT 122 a 72 c 54 g 118 t

ORIGIN

Alignment Scores:

Pred. No.: 164 Length: 366  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 6.84% Indels: 0  
DB: 17 Gaps: 0

US-09-897-438B-2 (1-117) x AZ791276 (1-366)

Qy 30 ThrThrAlaSerValLeuGln 37

Db 281 ACTACTACACCAAGTGTCTGCAG 258

RESULT 11

C47012/c

LOCUS

DEFINITION

C47012

Caenorhabditis elegans cDNA clone yk428h6 5', mRNA sequence.

ACCESSION C47012

VERSION C47012.1

KEYWORDS

SOURCE

ORGANISM

Caenorhabditis elegans.

Caenorhabditis elegans

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae

; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE

AUTHORS

1 (bases 1 to 371)

Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A., Sano

,M., Miyata,A. and Nishigaki,A.

Expression map of the C.elegans genome

Unpublished (1996)

CONTACT: Yuji Kohara

Genome Biology Lab.

National Institute of Genetics  
Yata 1111, Mishima, Shizuoka 411, Japan  
Tel: 81-559-81-6854  
Fax: 81-559-81-6855  
Email: ykoha@lab.nig.ac.jp.  
Location/Qualifiers

## FEATURES

source

1..371  
/organism="Caenorhabditis elegans"  
/strain="N2"  
/db\_xref="taxon:6239"  
/clone="yk428h6"  
/clone\_lib="Yuji Kohara unpublished cDNA:Strain N2  
hermaphrodite embryo"  
/sex="hermaphrodite"  
/dev\_stage="embryo"

BASE COUNT 97 a 71 c 80 g 123 t

## ORIGIN

## Alignment Scores:

Pred. No.: 166 Length: 371  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 6.84% Indels: 0  
DB: 14 Gaps: 0

US-09-897-438B-2 (1-117) x C47012 (1-371)

QY 43 GlySerCysArgPheSerTyrSer 50

Db 169 GGAAGTGTGCATTTTCATCTCT 146

## RESULT 12

AZ033696/c

LOCUS RPCI-23-248G7.TJ RPCI-23 Mus musculus genomic clone RPCI-23-248G7,  
DNA sequence.

ACCESSION AZ033696

VERSION AZ033696.1 GI:7118183

KEYWORDS GSS.

SOURCE house mouse.

## ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 381)  
Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret  
B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.  
and Fraser,C.M.  
Mouse BAC End Sequences from Library RPCI-23

## REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Other\_GSSs: RPCI-23-248G7.TV

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)  
or from Resea ch Genetics ([info@resgen.com](mailto:info@resgen.com)). BAC end page:  
[http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)  
Plate: 248 row: G column: 7  
Seq primer: SP6  
Class: BAC ends.

## FEATURES

source

1..381  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-23-248G7"

/clone\_lib="RPCI-23"  
/sex="Female"  
/lab\_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site\_1:  
ECORI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or  
brain genomic DNA was isolated and partially digested  
with a combination of EcoRI and EcoRI Methylase. Size  
selected DNA was cloned into the pBACe3.6 vector at the  
ECORI sites. The ligation products were transformed into  
DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 107 a 73 c 76 g 125 t

## ORIGIN

## Alignment Scores:

Pred. No.: 172 Length: 381  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 6.84% Indels: 0  
DB: 17 Gaps: 0

US-09-897-438B-2 (1-117) x AZ033696 (1-381)

QY 30 ThrThrAlaSerValLeuGln 37

Db 112 ACTACTACGCAAGTGTCTGCAG 89

## RESULT 13

AQ087955/c

LOCUS

DEFINITION HS\_2191\_B1\_D02\_MR CIT Approved Human Genomic Sperm Library D Homo  
sapiens genomic clone Plate=2191 Col=3 Row=H, DNA sequence.

ACCESSION AQ087955

VERSION AQ087955.1 GI:3456866

KEYWORDS GSS.

SOURCE human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome.  
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Sequence Tagged Connector

Plate: 2191 row: H column: 3

Class: BAC ends

High quality sequence stop: 381.

Location/Qualifiers

1..381

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="Plate=2191 Col=3 Row=H"

/clone\_lib="CIT Approved Human Genomic Sperm Library D"

/sex="male"

/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in

E-Coli DH10B"

BASE COUNT 69 a 81 c 131 g 100 t

## ORIGIN

## Alignment Scores:

Pred. No.: 172 Length: 381

Score: 8.00 Matches: 8

Percent Similarity: 100.00%  
 Best Local Similarity: 100.00%  
 Query Match: 6.84%  
 DB: 17

Conservative: 0  
 Mismatches: 0  
 Indels: 0  
 Gaps: 0

US-09-897-438B-2 (1-117) x AQ087955 (1-381)

Qy 46 ArgPheSerTyrSerAspProSer 53  
 Db 186 AGATTCTCCTACTCTGACCCCTCT 163

RESULT 14

AQ030475

LOCUS

DEFINITION HS\_2183.AL.G06.MF CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate-2183 Col-11 Row=M, DNA sequence.

ACCESSION AQ030475

VERSION AQ030475.1 GI:3275401

KEYWORDS GSS.

SOURCE human.

ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 399)  
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
 Hood,L.

Sequence-tagged connectors: A sequence approach to mapping and

scanning the human genome

Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

99380589

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Sequence Tagged Connector

Plate: 2183 row: M column: 11

Class: BAC ends

High quality sequence stop: 399.

FEATURES

source

1..399

Location/Qualifiers

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="plate-2183 Col-11 Row=M"

/notes="Organ: sperm; Vector: pBelobAC11; BAC Clones in

E-Coli DH10B"

BASE COUNT 114 a 100 c 57 g 123 t 5 others

ORIGIN

Alignment Scores:

Pred. No.: 183 Length: 399

Score: 8.00 Matches: 8

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 6.84% Indels: 0

DB: 17 Gaps: 0

US-09-897-438B-2 (1-117) x AQ030475 (1-399)

Qy 76 SerAsnValSerThrValIleHis 83

Db 245 TCAAATGTTTCACTGTCTATCCAT 268

RESULT 15

AW696758

LOCUS

DEFINITION NF108E09STIF1070 Developing stem Medicago truncatula cDNA clone

NF108E09ST 5', mRNA sequence.

400 bp mRNA linear EST 20-DEC-2000

ACCESSION

AW696758

VERSION AW696758.2 GI:11934078

KEYWORDS EST.

SOURCE barrel medic.

ORGANISM

Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
 Medicago.

1 (bases 1 to 400)

He,X.-Z., Shadle,G., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell  
 ,C.J., Flores,H.R., Inman,J.T., Weller,J.W., May,G.D. and Dixon  
 ,R.A.

TITLE

Expressed Sequence Tags from the Samuel Roberts Noble Foundation

Medicago truncatula stem library

Unpublished (2000)

JOURNAL

COMMENT

On Apr 14, 2000 this sequence version replaced gi:7571520.

Contact: Dixon RA

Plant Biology Division

The Samuel Roberts Noble Foundation

2510 Sam Noble Parkway, Ardmore, OK 73402, USA

Tel: 580 221 7302

Fax: 580 221 7380

Email: radixon@noble.org

Insert Length: 738 Std Error: 0.00

Plate: 108 row: E column: 09

Seq primer: TCACACAGGAACAGCTATGAC.

FEATURES

source

1..400

Location/Qualifiers

/organism="Medicago truncatula"

/db\_xref="taxon:3880"

/clone="NF108E09ST"

/clone\_lib="Developing stem"

/tissue\_type="stem"

/dev\_stage="Pooled developmental"

/notes="Vector: Lambda zap; Contains a mixture of

internodal stem segments"

BASE COUNT 93 a 87 c 105 g 115 t

ORIGIN

Alignment Scores:

Pred. No.: 184 Length: 400

Score: 8.00 Matches: 8

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 6.84% Indels: 0

DB: 10 Gaps: 0

US-09-897-438B-2 (1-117) x AW696758 (1-400)

Qy 74 AlaProSerAsnValSerThrVal 81

Db 312 GCTCCTCCCAATGTGTCAACGGTC 289

Search completed: November 6, 2002, 23:35:29

Job time : 2218 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 6, 2002, 18:13:22 ; Search time 2770 Seconds  
(without alignments)  
1229.252 Million cell updates/sec

Title: US-09-897-438B-2  
Perfect score: 625  
Sequence: 1 EQCGTIMHGNAVTFCEPYGP.....FQWKQDSLVRGVEYACWAL 117

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp  
-Q=/cgn2\_1/USFIO.spool/US09897438/runat\_06112002\_101924\_11289/app\_query.fasta\_1.263  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09897438.ecgn.1.1.2659 @runat\_06112002\_101924\_11289 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*

- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	625	100.0	11673	6	AX305377 Sequence
2	625	100.0	11673	10	U24703 Mus musculus
3	596	95.4	11187	10	AB049473 Rattus no
4	572	91.5	11580	6	AX410790 Sequence
5	572	91.5	11580	9	U79716 Human reell
6	572	91.5	11580	11	G30936 SWSS2926 Er
7	572	91.5	11580	11	G30938 SWSS3176 Er
8	418	66.9	10634	5	AF090441 Gallus ga
9	241	38.6	202764	2	AC128022 Rattus no
10	240	38.4	163985	10	AC121878 Mus muscu
11	240	38.4	183156	2	AC023062 Mus muscu
12	237	37.9	183641	2	AC124933 Rattus no
13	231	37.0	126130	2	AC095877 Rattus no
14	224	35.8	93163	9	HSAC000121 Human BAC
15	144	23.0	185996	2	AC041023 Homo sapi
16	118	18.9	158	4	AF232904 Bos tauru
17	92	14.7	4960	10	MUSBALBC1
18	89.5	14.3	1259	5	AF090951 Lacerata v
19	87.5	14.0	2092	5	AF090843 Emys orbi
20	86.5	13.8	231756	2	AC109258 Mus muscu
21	85.5	13.7	142100	2	AC094067 Rattus no
22	84	13.4	166764	2	AC129106 Homo sapi
23	84	13.4	170606	9	AC018898 Homo sapi
24	84	13.4	171862	2	AC009442 Homo sapi
25	84	13.4	343300	2	AC107296 Homo sapi
26	83	13.3	1845	6	I07862 Sequence 2
27	83	13.3	1845	6	I08324 Sequence 3
28	83	13.3	1845	6	I08327 Sequence 6
29	83	13.3	3408	6	I08074 Sequence 1
30	83	13.3	3408	6	I08323 Sequence 1
31	83	13.3	10519	1	AE001040 Archaeogl
32	81	13.0	2514	8	S73370 glucoamylas
33	81	13.0	2602	6	AR198404 Sequence
34	81	13.0	2602	6	AX073733 Sequence
35	81	13.0	2602	6	AX073733 Sequence
36	81	13.0	2602	6	ANGA01
37	81	13.0	3411	8	ASNGIGII
38	80.5	12.9	110000	2	LMFLCHR12_3
39	80.5	12.9	110000	2	LMFLCHR16_03
40	80	12.8	3411	6	E00315 DNA coding
41	80	12.8	4819	5	MGU49431
42	80	12.8	19823	2	AC019868 Drosophil
43	80	12.8	43525	3	U88311 Caenorhabd
44	80	12.8	163378	2	AC008336 Drosophil
45	80	12.8	174681	2	AC107156 Rattus no

ALIGNMENTS

RESULT 1

```
AX305377
LOCUS AX305377 11673 bp DNA linear PAT 11-DEC-2001
DEFINITION Sequence 128 from Patent WO0188188.
ACCESSION AX305377
VERSION AX305377.1 GI:17644926
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1
AUTHORS Ishikawa,K., Asai,S., Takahashi,Y., Nagata,T. and Ishii,Y.
TITLE Method for examining ischemic conditions
JOURNAL Patent: WO 0188188-A 128 22-NOV-2001;
SCHOOL Juridical Person Nihon University (JP)
FEATURES
source
1. .11673
/organism="Mus musculus"
/db_xref="taxon:10090"
BASE COUNT 2831 a 2985 c 2985 g 2872 t
ORIGIN
Alignment Scores:
Pred. No.: 6,42e-66 Length: 11673
Score: 625.00 Matches: 117
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-09-897-438B-2 (1-117) x AX305377 (1-11673)
QY 1 GluInCysGlyThrIleMetHisGlyAsnAlaValThrPheCysGluProTyrGlyPro 20
|||||
Db 970 GAGCAGTGGCCACCATCATGTCATGCAATGCTGTCACCTCTCTGAGCGCTAGCGCCCT 1029
QY 21 ArgGluLeuThrThrThrCysLeuAsnThrThrThrAlaSerValLeuGlnPheSerIle 40
|||||
Db 1030 CGAGAGCTGACCACACACATGCTGTAACACAAACAGACATGCTGCTCCAGTTTCCATT 1089
QY 41 GlySerGlySerCysArgPheSerTyrSerAspProSerIleThrValSerTyrAlaLys 60
|||||
Db 1090 GGTGAGATCATGTCGATTTAGTTACTCTGACCCACATCATCTGTCATACGCCAAG 1149
QY 61 AsnAsnThrAlaAspTrpIleGlnLeuGluLysIleArgAlaProSerAsnValSerThr 80
|||||
Db 1150 AACAAATACCGTGTGATTGAGTTCAGTGGAGAAATAGAGCCCTTCCAAATGTGAGCACA 1209
QY 81 ValIleHisIleLeuTyrLeuProGluGluAlaLysGlyGluSerValGlnPheGlnTrp 100
|||||
Db 1210 GTCAATCCATCCCTGATCTACCTCCCGAGGAGCAAGGAGAGCGTGCAGTTCAGTGG 1269
QY 101 LysGlnAspSerLeuArgValGlyGluValTyrGluAlaCysTrpAlaLeu 117
|||||
Db 1270 AACACAGACAGCGTGGAGTGGTGGGTGAGGTGTATGAGGCTGCTGGGCCCTG 1320
RESULT 2
MMU24703
LOCUS MMU24703 11673 bp mRNA linear ROD 19-DEC-1997
DEFINITION Mus musculus reelin mRNA, complete cds.
ACCESSION U24703
VERSION U24703.1 GI:2702252
KEYWORDS mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 11673)
D'Arcangelo,G., Miao,G.G., Chen,S.C., Soares,H.D., Morgan,J.I. and
Curran,T.
A protein related to extracellular matrix proteins deleted in the
mouse mutant reeler
JOURNAL Nature 374 (6524), 719-723 (1995)
```

```
95231649
7715726
2 (bases 1 to 11673)
Direct Submission
Submitted (10-APR-1995) Roche Institute of Molecular Biology, 340
Kingsland St., Nutley, NJ 07110, USA
3 (bases 1 to 11673)
D'Arcangelo,G.
Direct Submission
Submitted (19-DEC-1997) Dev. Neurobiology, St. Jude Childr. Res.
Hosp, 332 N. Lauderdale, Memphis, TN 38105, USA
Sequence update by submitter
On Dec 19, 1997 this sequence version replaced gi:902486.
Location/Qualifiers
1. .11673
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="5"
/map="D5Gmr1"
/tissue_type="cerebellum"
/dev_stage="adult"
1. .282
283..10668
/codon_start=1
/product="reelin"
/protein_id="AAB91599.1"
/db_xref="GI:2702253"
/translaton="MERGCAAPRALVLAVLLLLATLARAATGYVPRESFFFLCTHH
GELEGDEGEVLLSLHAGNPTYYVPGQYHYHTISTSTFFDGLLVTGLTSTSIQSS
QSHGGSAFGGMSDHFQNMCMVAVSHVSHLPTNLSFVIRDDPDSQVGLNPN
TATHRGQVIFKDALAOQLCEGAPTEATYSLAETHSDSVLRADPDSQVGLNPN
IWEKSCNCEGECQTIMHGNATFCPEYPRELTNTTNTTASVLFQSIGSGSFR
SYSDPSITVSYAKNNTADWOLEKIRAPSNVSVIHLYPEAKGSGVFOQWQDNL
RVEVYEACWALDNILVINSAREVLEDNLDVDTGNWLFEPGATVKHSCQSDNSI
YFNGESSEFNATTRVDLSTEDIQWSEGESQPTGNDILGAVVACGCTVESL
SLVFKDGERKLCPTPYMDTGTGNLREFYVMGGICDPGVSHENDIILYIKEGKEHI
ALDTLAYSYKVPSSLVSVINPELOTPATKFCRQKSHQGNRNWVAFDFHVLPLP
ALSTMIQFISNLGCGTHQPNQSVLEFSTNHNHSLHTECLPEACGPHLPSTV
YSSNTSGMNRITIPLANALTRDTIRWRGTGPIILGNMAIDNVYIGTSCLPFCGR
GQTRHCKCDPGFSGPACEMASQTFPMFISEFSGARLSYHNFYSGRAEVSFGG
VLASGALVFNKDGRLQITSFIDSSQSRLOFTLRGSKSVLSTCAPDQEGVLL
HYSDNGITWKLLEHYSYVNHPRISVELPDDARQFGIOFRMQVYHSSQGEDVNA
IDEVMTSVLFSISLDTNLVEVTSGLFYLGNOVQPCGHDMTLCFTGDSKLASSMR
YVETQSMQIGASYMIQFSLWGGCQKTYPHMDNQVKLEYSAHGLTWHLVQEECLPSM
PSCQETTSASIHASEFTQWRVTVPKQKTSWGTATFMSQSYTADSEWALDNIYI
GOCPNMCSGHGDGVCDCDQYQCTECHPEALPSTIMSDFENPSSSDMOEVI
GGVVKPECGCVVSSGSLYFSKAKRQLVSDLDPSWVDVFOFYVLIQGESNACNK
PDSREBGLIQQNNNGGSIQNHLLAEMFSDFSKPRFVYLELPAAGTKPTCFRFRKKPV
FSGEDYDQWAVDDIIILSEKQKQVIVVNPPTLPQNFYEKPAFDYPMQMSVWMLANE
GMAKNSFCATTPSAMVFGKSDGDRFAVTRDLTLKPGYVLIQKINIGCTSQFSSTAPV
LLQYSHDAGMSWFLLEKGCPPASAAKCGNSRELSEPTVYVYTGDFEWRITITAIPIR
SLASKTRFRWIOESSQKNVPPFGLDGVYISBPCSYCSGHGDCISGVCFLDGYTA
AOGTCVNTNHEMFDRFEGKSLPWYKITGGQVTCGCTLNDGRSLYNGLKREA
RTVPLDTRNISLVQFYLIQSKTSGITIIPRAYEGLVYQVYNDNGSLWHLLRELDF
MSLEPQIISLDLPREKATPATARFWQPKHSAQWALDVLGVNDSSTGTFQDK
LDGSDIQANNYRIQGGQVDIDCLMDLTALITENIGNPRIATFEDFVHVSSEFLQWE
MNGCKSPFSGAHIQIQLYSINNGKQWLTETECPTTIGCVHTSTSTYTSERFQW
RVVTVPLATNPTRFRWITQNTYVAGDSWADINVLASGCPWMSGRIGICDGLGRC
VCDRGGGCPVVPVPLPILKDDFNGLHPLDWPVYVGAERGNLNGETIKSGTCLIF
KGEGLMLISRLDLCNTMYVQPSLRFAKGTPEPERSHILQPSVSGCVTWHLMDEFF
FPQTTSLFINVPLPYGAQTNATFRWLQYNNKKEEIIWIDIFIIDGNLNNPVLL
LDTDFGPREDNMFYFPGNIGLYCYSSKGAPEEDSAMVFSNEVEHSTITRDLSV
NENTIIQFENVGCSTDSADPVRLEFDFGATWHLPLCYHSSLSVSSLCSTEH
HPSTYVAGTQWRREVHFHGLHLCGSVRFWYQGYFAGSQPVTAIDNVYIGQ
CEMCKYGHGSCINGTKICDPYSGPTCKISTKNPDFLKDDEGOLSEDRFLMSGCK
PSRCKGLSGNLLFNEDEGLMLVTRDLDLSHARFVQFRWLRCCKGKVPDPSPQVPL
LOYSNLGGLSSLLQLEFLFSSNVNGVRIALEMPLKARSKRNLRWQWQWSENGHFI
WVTDIQLIGNISGNTVEDDFTLDSRKLHHPGKTMPCVCGSTGALVFIERAKR
SYVHLORILVSDTFNKWTRITLPLPSYTRSOATFRWHQPAKQOTWADNVYIGD
GCLDMCSHGRCVQSGCDEOMGLYCDPEPISLTQLKDNFNRAFSNOMLTVSGS
KLSTVCGAVASGLALHFSGGCSRLITVTLNLNNAEFIQFYFMYGCLITPSNRNQGLV
```



HEATTYNAVNSWKKRITIQLPDHVSSATQFRWIOKGETEKOSWAIDHVTGEACP  
RLCSGHGYCTGTGAVCTCDESGDQCSVFSHELPSYIKONFESAYTEANWETIOGGA  
IGSCGGLAPYAGHDSLYFNGCQIRQAATKPLDLTRASKIMFVLQIGSTAOOTCSND  
LSPHNTVDKAVLLQYSVNNIGITWHVIAQHOPKDFTOAORVSYNVPLEARMKGVLLRW  
QPHNTEGHQDOWALDHVEVVLVSTRKYNMNFSRQHGLAHFNRRRSLRRYP"

BASE COUNT 2645 a 3029 c 2930 g 2583 t  
ORIGIN

Alignment Scores:  
Pred. No.: 2,25e-62 Length: 11187  
Score: 596.00 Matches: 111  
Percent Similarity: 97.44% Conservative: 3  
Best Local Similarity: 94.87% Mismatches: 3  
Query Match: 95.36% Indels: 0  
DB: 10 Gaps: 0

US-09-897-438B-2 (1-117) x AB049473 (1-11187)

Qy 1 GluGlnCysGlyThrIleMetHisGlyAsnAlaValThrPheCysGluProTyrGlyPro 20  
|||||  
Db 1031 GAACAGTGTGGCACCATCATGCATGGCAACGCTGCACCTTCTGTGAGCCATATGGTCT 1090

Qy 21 ArgGluLeuThrThrCysLeuAsnThrThrAlaSerValLeuGlnPheSerIle 40  
|||||  
Db 1091 CGAGATTTGACTACCATACCTGAACACGACAAACAGCATCCGCTCCAGTTCTCCATT 1150

Qy 41 GlySerGlySerCysArgPheSerTyrSerAspProSerIleThrValSerTyrAlaLys 60  
|||||  
Db 1151 GGGTCAGGATCTCGATTTAGTACTCTGACCCAGCATCATTTGTATATGCCAAG 1210

Qy 61 AsnAsnThrAlaAspTrpIleGlnLeuGluLysIleArgAlaProSerAsnValSerThr 80  
|||||  
Db 1211 AACAACTACTGCGGATGGATTGAGTGGAGAGATTAGAGCCCTTCCAAGTAAGCACC 1270

Qy 81 ValIleHisIleLeuTyrLeuProGluGluAlaLysGlyGluSerValGlnPheGlnTrp 100  
|||||  
Db 1271 ATCATCCACATCCCTACCTCTCCCTGAGGAGCCCAAGGGGAGACGTGCGAGTTCCAGTGG 1330

Qy 101 LysGlnAspSerLeuArgValGlyGluValTyrGluAlaCysTrpAlaLeu 117  
|||||  
Db 1331 AACAGGACAGCCTGCATGTGGTGAGGTGTACGAAGCCTGCTGGGCCCTG 1381

RESULT 4  
AX410790  
LOCUS AX410790 11580 bp DNA linear PAT 14-JUN-2002  
DEFINITION Sequence 3437 from Patent WO0229103.  
ACCESSION AX410790  
VERSION AX410790.1 GI:21443495  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1  
AUTHORS Alvares,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.  
TITLE Gene expression profiles in liver cancer  
JOURNAL Patent: WO 0229103-A 3437 11-APR-2002;  
GENE LOGIC INC (US)

FEATURES  
source Location/Qualifiers  
1..11580  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/note="EMBL/GenBank Accession No. U79716"  
BASE COUNT 3014 a 2696 c 2753 g 3116 t 1 others  
ORIGIN

Alignment Scores:  
Pred. No.: 2.1e-59 Length: 11580  
Score: 572.00 Matches: 106  
Percent Similarity: 95.73% Conservative: 6  
Best Local Similarity: 90.60% Mismatches: 5  
Query Match: 91.52% Indels: 0  
DB: 6 Gaps: 0

US-09-897-438B-2 (1-117) x AX410790 (1-11580)

Qy 1 GluGlnCysGlyThrIleMetHisGlyAsnAlaValThrPheCysGluProTyrGlyPro 20  
|||||  
Db 860 GAACAGTGTGGCGGATATGATGCATGGCAATGCCGTCACTTCTGTGAACCATATGGCCCA 919

Qy 21 ArgGluLeuThrThrCysLeuAsnThrThrAlaSerValLeuGlnPheSerIle 40  
|||||  
Db 920 CGAAGCTGATTACACACAGGCCCTTAATACACACACAGCTTCTGCTCCCAATTTCCATT 979

Qy 41 GlySerGlySerCysArgPheSerTyrSerAspProSerIleThrValSerTyrAlaLys 60  
|||||  
Db 980 GGGTCAGGTTTCATCGCTTAGTTATTACAGACCCAGCATCATCGTGTATATGCCAAG 1039

Qy 61 AsnAsnThrAlaAspTrpIleGlnLeuGluLysIleArgAlaProSerAsnValSerThr 80  
|||||  
Db 1040 AATAACCTGCGGACTGGATTTCAGCTAGAGAAATAGAGCCCTTCCAATGTGAGCACA 1099

Qy 81 ValIleHisIleLeuTyrLeuProGluGluAlaLysGlyGluSerValGlnPheGlnTrp 100  
|||||  
Db 1100 ATCATCCATCTCTACCTTCTCTGAGGAGCCCAAGGGGAGATGTCCCAATTTTCAGTGG 1159

Qy 101 LysGlnAspSerLeuArgValGlyGluValTyrGluAlaCysTrpAlaLeu 117  
|||||  
Db 1160 AAGCAGGAAATCTTCGTAGTGAAGTGTATGAAGCCTGCTGGGCTTA 1210

RESULT 5  
HSU79716  
LOCUS HSU79716 11580 bp mRNA linear PRI 25-FEB-1997  
DEFINITION Human reelin (RELN) mRNA, complete cds.  
ACCESSION U79716  
VERSION U79716.1 GI:1743884  
KEYWORDS Homo sapiens.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 11580)  
AUTHORS Desilva,U., D'Arcangelo,G., Braden,V.V., Chen,J., Miao,G.G.,  
Curran,T. and Green,E.D.  
TITLE The human reelin gene: isolation, sequencing, and mapping on  
chromosome 7  
JOURNAL Genome Res. 7 (2), 157-164 (1997)  
MEDLINE 97202106  
PUBMED 9049633

REFERENCE 2 (bases 1 to 11580)  
AUTHORS Desilva,U., D'Arcangelo,G., Braden,V.V., Chen,J., Miao,G.G.,  
Curran,T. and Green,E.D.  
TITLE Direct Submission  
JOURNAL Submitted (26-NOV-1996) National Center for Human Genome Research,  
National Institutes of Health, 49 Convent Drive, MSC4431, Bethesda,  
MD 20892, USA

FEATURES  
source Location/Qualifiers  
1..11580  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="7"  
1..11580  
/gene="RELN"  
176..10358  
/gene="RELN"  
/codon\_start=1  
/product="reelin"  
/protein\_id="AAC51105.1"  
/db\_xref="GI:1743885"  
/translation="MERSGWARTFLALLGLATLRAAAGYYPFRSPFFLCTHHG  
ELEGGDEQGEVLISLHAGNPTVYVPGQYHVTISTFDFGLLVGLYTSVQASQ  
SIGGSAFGFIMSDHQFNGFMCVSVASHSLPTNLNFIAPAGTGCYVFNWAT  
ATHRGQVIFKDALAQOLCEQAPDVTVPHLAEIHSDSILLRDFDSYHQLNPN  
WYECNCCETGEQCGATMHGNVATFCBPYGPRLITGLNTTASVLFQISGSSCRFS  
YSDPSITVLYAKNNSADWOLEKIRAPSNVSTIIHILYLPEDGAKGVNQFQWQENLR  
VGEVYEACWALNDNLIITNSAHQVWLEDSLDPVDTGNWULFFPGATVKHSCQSDGNSIY

FIGNESEFNFAATRDVLDLSTEDIQOWSEEFESQPTGWDVLGAVICTBCGTIESGLS  
 MYFLKDEKRLKTPSMDTGTGYNLREYFVYMGGICDPNGNSHENDIILYAEIKRGKHEIT  
 LDTLSYSSVPSVLSVSVINPELOTPATKFKLOKHQGNRRNVAWDFEHLVPLPS  
 TWSHMTQFSINLGGCTHQPONSVSLEESNHNRSWLLTECLPELCAQPHLPSTHYV  
 SENYSGWNRITIPALNALTNRIRWQRTGPIILGNMWAIDNVIYIGSCLKPCSGRG  
 QCTRHCCKDPGSPGACEMASQTFPMFISFSGSSRLSSYHNFSYIAEYSGFGGV  
 LASGKALVNEKGRRLITISFLDSQSRELOFTLRGSKSVLSTCRAPQPOQREGVLLH  
 YSDNGITWKLLEHYSYLSHEPRLIISVELPGDKQOFQIFRWQPVYHSQOREDWAI  
 DELIMTSVNSISDFTNLVETVQSLGFLGNVQPCYGHDMTLCPTGDSKLASSRMY  
 VETOSMOIGASVMIQPSLYMGCGKYTPHNDQKLEYSTNHLGTHWHLVQBECLPSMP  
 SQCFETSASIHASEFTQNRVILVLPQKTSWATFRWSSQTSYITADDEWALDSYIIG  
 QQCPCNSHGSGDHGICDQGYQGTCECHPEALPSTIMSDFENQWESQWQEVIG  
 GETVPCQCGVSSGSLYFKAGKRLQVSWDLDTSWDFVQFYIQIGESASCKNP  
 DSREEGVLQYNNNGIOWHLLAEYFDSKPRFYVLELPAAKPQCTFRFWMQVLF  
 SEDYDQWAVDDIILSEKOKIIPVINTLPQNEVEKPAFYPMNQMSVYMLQWLANEG  
 MYKNETFECAATPSAMIFGKSDGRFAVTRDILKPGYVLOFKLNGICANQFSTAPVL  
 LOYSHDAGMSWFLYKEGYPASAGCGCNSRELSEPTMTHTGDFPEWTRITIVIPRS  
 LASSKTRFWIOESSQKNVPFPLDGVYISEPCSYCSGHGDCISGVCDLGLGYTAA  
 QGTCVSNPNHNMEDRFEGKLSPIWYKITGAQVGTGCTLNDKSLYFNGPKREAR  
 TVPLDTRNIRLVQFYIQIGSKTSGITKIPRTNEGLIVQYSDNDGILWHLRLDFEM  
 STLEPOIISIDLPODAKTATAPRWQPOHKSQAOWALDDVLIGHNDSQSTCFQDFM  
 DSGIDLOANNVRIQGOVDIDCLSDMTALIFTENIGKPYAETWDSVYSASFQLEM  
 SMGCKPFSNSUSVQLOYSNLNGKWHLUTECVPTICGLHTTESYISERFQNMK  
 KITVILPLSTISPTFRFIQANYTVGADSWALDNVYLAGCPWMSGRCIGDAGRCV  
 CDRGFGGVCYVPVPLPSILKDDFNGLHPDLMEYVYGAERNGLNGETIKSGTSLIFK  
 GEGRLMLISRDLDCTNMTYQFSLRFIAKSTPERSHILQFISGGITWHLMDPEYF  
 POTNTLFTNVPLPYTAQTNATFRWQPYNNCKKEIWIIVDDFIIDGNWNNVPMLL  
 DTFQEGPREDNFFYPGGNIGLYCYPSSKGAPEEDSAMFVSNVEGHSITTRDLNVN  
 ENTIIQFELVNGCSTSSADPVRLEFSRDFGATWHLPLCTHSSSHVSLCSTDEHH  
 PSTTYQFNGQWRREVHFGKLHLCGSVFRWYQGFYPAGSQPVTAIDNVIYIGPC  
 EEMCGQSGINGTKICDQPSGPTKISRNPDFLKDQFEGQLESDFRLMISGSGKP  
 SRKCGILSGNLLFFNEDGLRMLTRDLDSHARFQFWMRLGCGVDPDRSQPVLL  
 QYSLNGLSNLLQEPFLSNNGVRYIALEIPLKARSSTRLRWQPSSENGHFYSWP  
 VIDLIIGNISGNTVLEDFTLDSRWLLHPGKMPVCGSTGDALVPIEKASTRY  
 VYSTDAVNEDEFQIDFAASVSDSCYAIETYSVDLSLWHPVROCLPTNVNCS  
 RYHLQRLVSDTFNKWTRITLPLPPYRQATFRHQPAPFDKQOTWADNVIYIGDG  
 CHLQMSGHRCIQGNCVQDEQWGLYCDPSETSLPTQLKDNFNAPQSSWMLTVNGGK  
 LSTVGAVASGMLHMEIFYDQYSKPGFVNILLPPDAKEIATFRFWMQPRHGDNDWA  
 EYSVNGITWMLLEIFYDQYSKPGFVNILLPPDAKEIATFRFWMQPRHGDNDWA  
 IDNVLISGSADQRTVMDLTFESSAPVPOHRSADAGPVGRIADFMDPKTSVNEHL  
 FIDDCITVERCSDPGVMLCGSHDREYVATVHDLTPTEGIMQFKISVCGKXSEKIA  
 QNCHVQVSTDFGVSNVLYPOCLPADKPSGVSPQSVFFFTPKMKRITYPLPESLY  
 GNPRFRFYKIDMGOWAIDNFIYLGPCLDNCGHGDREQCICDPGSGNRCYLTTH  
 TLKTLFKERFDEEIKPDLWMSLEGSTCEGILAEADTALYFGSSTVQVQATDLDL  
 RGAFLQYWGVRIGSENMTSCHRPICRKEGVLLDYTDGDTITWLLHEMDYOKYISVR  
 HDYILLPEDALNTTTLRWQPVVINSVGSVERAQWALNILLGAEIINPSQLVD  
 TPEDEGTSHEENWSFYPNAVTRAGFCNFSFLHWPNNKKDKTHNALSSRELIIQGY  
 MMQFKIVVQCEATSCGDLHSMLEYTKDARSQVQVQCLPSSNSIGCSPPQFHE  
 ATIYNSVNSWKRRITIQPDHVSSATQFRWIKGEETEKSQWAIHDHYIYIGACPKL  
 CSHGQYCTTGATCICDESFGQDCSVFSDHLPYIKDNFESARTANWETIQGVIG  
 SCGGOLAPYAHGDSLYFNGCQIROAATKPLDLTRASKIMFVLQIGSMQSDSCNSDL  
 GPHAVDKAVILQYSVNNGITWVIAHQPKDFTQAOQVSYNVPLEARMKGVLLRWQPV  
 RINGTHQDQWALDHVBVILVSTKQNMNFSQKHLRHFYNNRRSLRRYP"

BASE COUNT 3014 a 2696 c 2753 g 3116 t 1 others

ORIGIN

Alignment Scores:

Pred. No.: 2,1e-59 Length: 11580  
 Score: 572.00 Matches: 106  
 Percent Similarity: 95.73% Conservative: 6  
 Best Local Similarity: 90.60% Mismatches: 5  
 Query Match: 91.52% Indels: 0  
 DB: Gaps: 0

US-09-897-438b-2 (1-117) x HSU79716 (1-11580)

QY 1 GluGlnCysGlyThrIleMetHisGlyAsnAlaValThrPheCysGluProTyroGlyPro 20  
 Db 860 GAACAGTGTGGCGGATATGATGCGAATGCCGTACCTTCTGTGAACCATATGGCCCA 919  
 QY 21 ArgGluLeuThrThrCysLeuAsnThrThrThraAlaSerValLeuGlnPheSerIle 40  
 Db 920 CGAGAACTGATTACCAACAGCCCTTAATACAAACACAGCTTCTGTCTCCCAATTTTCATT 979

QY 41 GlySerGlySerCysArgPheSerTySerAspProSerIleThrValSerTyAlaLys 60  
 Db 980 GGGTCAGTTTCATGTGCTCTTTAGTTATTCAGACCCAGCATCATCGTGTATATATGCCAAG 1039  
 QY 61 AsnAsnThrAlaAspTrpIleGlnLeuGluLysIleArgAlaProSerAsnValSerThr 80  
 Db 1040 AATAACTCTCGGACTGGAITCAGTAGAGAAAATTAGAGCCCTTCCCAATGTCAGCACA 1099  
 QY 81 ValIleHisIleLeuTyLeuProGluGluAlaLysGlyGluSerValGlnPheGlnTrp 100  
 Db 1100 ATCATCATATCTCTACCTTCCTGAGAGCCCAAGGAGAGATGTCCAATTTCACTAGTGG 1159  
 QY 101 LysGlnAspSerLeuArgValGlyGluValTyrcGluAlaCysTrpAlaLeu 117  
 Db 1160 AAGCAGGAAAATCTTCGTAGGTGAAGTGTATGAAGCCCTGCTGGGCCTTA 1210  
 RESULT 6  
 G30936 11580 bp DNA linear STS 28-SEP-1998  
 LOCUS SHSS2926 Eric D. Green Homo sapiens STS genomic, sequence tagged  
 DEFINITION G30936  
 ACCESSION G30936  
 VERSION G30936.1 GI:19233209  
 KEYWORDS STS.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 Bouffard, G.G., Iyer, L.M., Idol, J.R., Braden, V.V., Cunningham, A.F.,  
 Weintraub, L.A., Mohr-Tidwell, R.M., Peluso, D.C., Fulton, R.S.,  
 Leckie, M.P. and Green, E.D.  
 A collection of 1814 human chromosome 7-specific STSs  
 Genome Res. 7 (1), 59-64 (1997)  
 97189344  
 9037602  
 REFERENCE 2 (bases 1 to 11580)  
 Human chromosome 7 STSs (1997)  
 Unpublished (1997)  
 On Apr 3, 1997 this sequence version replaced gi:1706935.  
 Synonyms: RELN  
 GDB: DSEG: RELN  
 Contact: Eric D. Green  
 Genome Technology Branch  
 National Human Genome Research Institute/NIH  
 49 Convent Dr., MSC4431, Bldg. 49, Rm. 2A08, Bethesda, MD 20892  
 Tel: 3014020201  
 Fax: 3014024735  
 Email: egreen@hgrl.nih.gov  
 Primer A: CTCCTACCTTCCTGAGAGCGCCAA  
 Primer B: AAGCCAGTTCCTGTGTCCACTG  
 STS size: 183  
 PCR Profile:  
 Presoak: 0 degrees C for 0.00 minute(s)  
 Denaturation: 92 degrees C for 0.17 minute(s)  
 Annealing: 55 degrees C for 1.00 minute(s)  
 Polymerization: 72 degrees C for 1.00 minute(s)  
 PCR Cycles: 35  
 Thermal Cycler: PerkinElmer 9600  
 Protocol:  
 Template: 30-100 ng  
 Primer: each 1 uM  
 dNTPs: each 200 uM  
 Taq Polymerase: 0.05 units/ul  
 Total Vol: 10 ul  
 Buffer:  
 MgCl2: 1.5 mM  
 KCl: 100 mM  
 Tris-HCl: 10 mM  
 NH4Cl: 5 mM

pH: 8.6

This STS was developed from sequence determined by another investigator. See GenBank record: U79716 For additional information about the NHGRI chromosome 7 mapping project, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>. Also see Genomics 11:548-64 (1991) [MUID-92128937].

## FEATURES

source 1..11580  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="7"  
/clone\_lib="Eric D. Green"  
1..11580  
/gene="RELN"  
1112..1294  
/gene="RELN"  
1112..1134  
/gene="RELN"  
primer\_bind 3014 a 2696 c 2753 g 3116 t 1 others  
BASE COUNT 3014 a 2696 c 2753 g 3116 t  
ORIGIN

Alignment Scores:  
Pred. No.: 2,1e-59 Length: 11580  
Score: 572.00 Matches: 106  
Percent Similarity: 95.73% Conservative: 6  
Best Local Similarity: 90.60% Mismatches: 5  
Query Match: 91.52% Indels: 0  
DB: 11 Gaps: 0

US-09-897-438B-2 (1-117) x G30936 (1-11580)

Qy 1 GluGlnCysGlyThrIleMetHisGlyAsnAlaValThrPheCysGluProTyrGlyPro 20  
Db 860 GAACAGTGTGGCGGATTCATCGTGGCAATCCGCTCTGTGAACCATATGGCCCA 919  
Qy 21 ArgGluLeuThrThrCysLeuAsnThrThrThrAlaSerValLeuGlnPheSerIle 40  
Db 920 CGAGACTGATTACCAAGCGCTTAATACACACAGCTTCTGTCCTCCATTTCCATT 979  
Qy 41 GlySerGlySerCysArgPheSerTyrSerAspProSerIleThrValSerTyrAlaLys 60  
Db 980 GGGTCAGGTTTCATGTCGCTTTAGTATTATTCAGACCCAGCATCATCGTGTATGCCAAG 1039  
Qy 61 AsnAsnThrAlaAspTrpIleGlnLeuGluLysIleArgAlaProSerAsnValSerThr 80  
Db 1040 AATACCTCGCGGACTGATTCAGTAGAGAAATAGAGCCCTTCCAAATGTCAGCACA 1099  
Qy 81 ValIleHisIleLeuTyrLeuProGluAlaLysGlyGluSerValGlnPheGlnTrp 100  
Db 1100 ATCATCCATATCTCTACCTTCTGTGAGGACGCCAAGGGAGAAATGTCCAATTTTCAGTGG 1159  
Qy 101 LysGlnAspSerLeuArgValGlyGluValTyrGluAlaCysTrpAlaLeu 117  
Db 1160 AAGCAGGAAAAATCTCGTGTAGTGAAGTGTATGAAGCCTCTGGGCGCTTA 1210

## RESULT 7

G30938 G30938 11580 bp DNA linear STS 28-SEP-1998  
LOCUS SWSS3176 Eric D. Green Homo sapiens STS genomic, sequence tagged  
DEFINITION site.

ACCESSION G30938

VERSION G30938.1 GI:1923211

KEYWORDS STS.

SOURCE Homo sapiens.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 11580)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS Bouffard,G.G., Iyer,L.M., Idol,J.R., Braden,V.V., Cunningham,A.F.,  
Weintraub,L.A., Mohr-Tidwell,R.M., Peluso,D.C., Fulton,R.S.,  
Leckie,M.P. and Green,E.D.

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

A collection of 1814 human chromosome 7-specific STS  
Genome Res. 7 (1), 59-64 (1997)  
97189344  
9037602  
2 (bases 1 to 11580)  
Green,E.D.  
Human chromosome 7 STS (1997)  
Unpublished (1997)  
On Apr 3, 1997 this sequence version replaced gi:1706937.  
Synonyms: RELN  
GDB\_DSEG: RELN  
Contact: Eric D. Green  
Genome Technology Branch  
National Human Genome Research Institute/NIH  
49 Convent Dr., MSC4431, Bldg. 49, Rm. 2A08, Bethesda, MD 20892  
Tel: 3014020201  
Fax: 3014024735  
Email: egreen@nhgri.nih.gov  
Primer A: TGTGGCTGATTTACTAACAG  
Primer B: TGGGCTTTTATAGCTTG  
STS size: 238  
PCR Profile:

Presoak: 0 degrees C for 0.00 minute(s)  
Denaturation: 92 degrees C for 0.17 minute(s)  
Annealing: 50 degrees C for 1.00 minute(s)  
Polymerization: 72 degrees C for 1.00 minute(s)  
PCR Cycles: 35  
Thermal Cycler: PerkinElmer 9600

Protocol:  
Template: 30-100 ng  
Primer: each 1 uM  
dNTPs: each 200 uM  
Taq Polymerase: 0.05 units/ul  
Total Vol: 10 ul

Buffer:  
MgCl2: 1.5 mM  
KCl: 100 mM  
Tris-HCl: 10 mM  
NH4Cl: 5 mM  
pH: 8.6

This STS was developed from sequence determined by another investigator. See GenBank record: U79716 For additional information about the NHGRI chromosome 7 mapping project, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>. Also see Genomics 11:548-64 (1991) [MUID-92128937].

## FEATURES

source 1..11580  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="7"  
/clone\_lib="Eric D. Green"  
1..11580  
/gene="RELN"  
11314..11551  
/gene="RELN"  
11314..11333  
/gene="RELN"  
primer\_bind 3014 a 2696 c 2753 g 3116 t 1 others  
BASE COUNT 3014 a 2696 c 2753 g 3116 t  
ORIGIN

## Alignment Scores:

Pred. No.: 2,1e-59 Length: 11580  
Score: 572.00 Matches: 106  
Percent Similarity: 95.73% Conservative: 6  
Best Local Similarity: 90.60% Mismatches: 5  
Query Match: 91.52% Indels: 0  
DB: 11 Gaps: 0

US-09-897-438B-2 (1-117) x G30938 (1-11580)

Qy	1	GlulnGlyThrIleMetHisGlyAsnAlaValThrPhcCysGluProTyrGlyPro	20
Db	860	GAACAGTGTGCGCGATATATCATGTGGCAATGCCGTACCTTCTGTGAACCATATGGCCCA	919
Qy	21	ArgGluLeuThrThrCysLeuAsnThrThrThrAlaSerValLeuGlnPhSerIle	40
Db	920	CGAGACTGATTACACAGCGCTTATATACACACAGCTTCTGTCTCCAATTTTCCATT	979
Qy	41	GlySerGlySerCysArgPhSerTyrSerAspProSerIleThrValSerTyrAlaLys	60
Db	980	GGGTGAGTTCATGTCGCTTTAGTTATTTCAGACCCAGCATCATCGTGTATATGCCAAG	1039
Qy	61	AsnAsnThrAlaAspTrrIleGlnLeuGluLysIleArgAlaProSerAsnValSerThr	80
Db	1040	AATAACTTCGCGGACTGGATTACAGTACAGAGAAATATAGAGCCCTTCCAATGTGAGCACA	1099
Qy	81	ValIleHisIleLeuTyrLeuProGluGluAlaLysGlyGluSerValGlnPhSerIleTrrp	100
Db	1100	ATCATCATATCTCTACCTTCTGAGGACGCCAAGGGGAGAAATGTCCAATTTTCAGTGG	1159
Qy	101	LysGlnAspSerLeuArgValGlyGluValTyrGluAlaCysTrrpAlaLeu	117
Db	1160	AAGCAGGAAATCTTCGTGTAGTGAAGTGTATGAAGCCTGCTGGGCCTTA	1210
RESULT	8		
AF090441			
LOCUS		10634 bp	mRNA linear VRT 15-SEP-1998
DEFINITION		Gallus gallus extracellular reelin (Rein) mRNA, partial cds.	
ACCESSION		AF090441	
VERSION		AF090441.1	GI:3600101
KEYWORDS			
SOURCE		Gallus gallus.	
ORGANISM		Gallus gallus	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS		Archosauria; Aves; Neognathae; Galliformes; Phasianidae;	
TITLE		Phasianinae; Gallus.	
JOURNAL		1 (bases 1 to 10634)	
REFERENCE		Bernier B. and Goffinet, A.M.	
AUTHORS		2 (bases 1 to 10634)	
TITLE		Submitted (04-SEP-1998) Neurobiology, Fac.N-D de la Paix, Rue de	
JOURNAL		Bruxelles, 61, Namur B-5000, Belgium	
FEATURES		Location/Qualifiers	
source		1..10634	
gene		/organism="Gallus gallus"	
		/db_xref="taxon:9031"	
		<1..10634	
		/gene="Rein"	
		<1..9631	
CDS		/gene="Rein"	
		/codon_start=2	
		/product="extracellular reelin"	
		/protein_id="AAC35559.1"	
		/db_xref="GI:3600102"	
		/translations="TGLNTTASVLQFSLGSGCRFSYSDPSITVSYSKNSSADWTO	
		LEKISAPSNVSTIIHLYLPEDAKGENVHFQKQDYLHAGEVEACWALNILLINAA	
		HRKVLENDLPDVTGKMLFPFPAQVHKSCSDGNSIYFHGTGEGSEFNKATTRDVLDS	
		TEDAQBQWAEFSPQKQDILGAVIGTEGTLSESSSMVFLRGRGKTCPTMYDTTG	
		YGNLQYTFMGMGNCDSGESHENDVILYAKTEGRREHIALDLTYAYKVPVLSVVIS	
		PDLQTPATFCLKOKSHOGRNRNWAVDYFHLVPLPSTVTHTMIOFSNLGCGCTYOP	
		NSVLSFEFTNHGRSWSLHTECLPEICACGPHLPSTVYASENSGNRITTPVPAAL	
		TSOTRRWRGTGPIHGNMAIDNIYIGPSCLKFCSGRGQCTRGKCDPGFSGPACET	
		ASOTPFMFISESPASRLSSYHNFYISIRGAEVFGCVGLASGALVFNKDRGLTAL	
		FLDSOSSRFLQTLRLGSKSVLSTCKAPQDPQEGVLLHYSDNGITWKLLEHYSLYN	
		HEPRTISVELPEDARQIGIQRWMPQVHSSQGEDVMAIDIVMTSVLFSNLISLDFTNL	
		VEYTSQSLGVLGNVPYCGSHDWTLCFTGDSKLTSSMRYVETQSMOIGASYMIOFLNM	
		GCCOKETPHMDNQKLEYSTNHGILTWHVQECPLSPMSPQCEFTASIIYHSNEFTQWR	
		RITVLLPQKTWSATRFNWSQCIYAPDEWALNDIYIGQCPNMCSGHGWDHGVCVCH	
		DSFGRTECPENPLTSRPNDFVLTKEWQIIGIEIVKPEEGCGVIGSSGSLY	
		FNKAGRQLVSWPDLTWDVDFVFIQIGESSSCNRPDSSREBGLVLYQSNNGGINQW	

RESULT 9  
AC128022/c  
LOCUS  
DEFINITION Rattus norvegicus clone CH230-525L20, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 97 unordered pieces.  
AC128022  
AC128022.1 GI:21908605  
VERSION HTG; HTGS-PHASE1.  
KEYWORDS Rattus norvegicus  
SOURCE Rattus norvegicus  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 202764)  
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,  
Alsbrooks, S.L., Amarantunga, H.C., Are, J.R., Ayele, M., Banks, T.,  
Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,  
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,  
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,  
Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,  
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,  
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,  
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
Correll, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,  
Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B.,  
Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,  
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,  
Karlssoon, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,  
Li, J., Li, Z., Lichtharge, O., Lieu, C., Liu, J., Liu, W., Louissegh, H.,  
Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,  
Massey, E., Mahoney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,  
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
Nguyen, N., Nickerson, E., Nwokenwo, S., Oguh, M., Okwunodu, G.,  
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
Peters, L., Picken, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,  
Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,  
Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I.,  
Sodergren, E., Sonake, T., Sparks, A., Stanley, H., Stone, H.,  
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
Usmani, K., Vasquez, L., Vera, Y., Villalon, D., Vinson, R., Wang, Q.,  
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,  
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
Weinstock, G. and Gibbs, R.  
Direct Submission  
Unpublished  
2 (bases 1 to 202764)  
Worley, K.C.  
Direct Submission  
Submitted (19-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: KAIV  
Center clone name: CH230-525L20  
----- Summary Statistics  
Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye: 100% of reads

Assembly program: Phrap; version 0.990329  
Consensus quality: 113069 bases at least Q40  
Consensus quality: 119970 bases at least Q30  
Consensus quality: 125656 bases at least Q20  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 97 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 1244: contig of 1244 bp in length  
\* 1245 1344: gap of unknown length  
\* 1345 2346: contig of 1002 bp in length  
\* 2347 2447: gap of unknown length  
\* 2448 3813: contig of 1367 bp in length  
\* 3814 3913: gap of unknown length  
\* 3914 5101: contig of 1188 bp in length  
\* 5102 5201: gap of unknown length  
\* 5202 6824: contig of 1523 bp in length  
\* 6825 8048: contig of 1224 bp in length  
\* 8049 9212: contig of 1064 bp in length  
\* 9213 9312: gap of unknown length  
\* 9313 10432: contig of 1120 bp in length  
\* 10433 10532: gap of unknown length  
\* 10533 11890: contig of 1358 bp in length  
\* 11891 11990: gap of unknown length  
\* 11991 13068: contig of 1078 bp in length  
\* 13069 13169: gap of unknown length  
\* 13170 14486: contig of 1318 bp in length  
\* 14487 14586: gap of unknown length  
\* 14587 15996: contig of 1410 bp in length  
\* 15997 16096: gap of unknown length  
\* 16097 17534: contig of 1438 bp in length  
\* 17535 17634: gap of unknown length  
\* 17635 18703: contig of 1069 bp in length  
\* 18704 18803: gap of unknown length  
\* 18804 19884: contig of 1081 bp in length  
\* 19885 19984: gap of unknown length  
\* 19985 21539: contig of 1555 bp in length  
\* 21540 21639: gap of unknown length  
\* 21640 22780: contig of 1141 bp in length  
\* 22781 22880: gap of unknown length  
\* 22881 23926: contig of 1046 bp in length  
\* 23927 24026: gap of unknown length  
\* 24027 25544: contig of 1518 bp in length  
\* 25545 25644: gap of unknown length  
\* 25645 26899: contig of 1255 bp in length  
\* 26900 26999: gap of unknown length  
\* 27000 28237: contig of 1238 bp in length  
\* 28238 28337: gap of unknown length  
\* 28338 29758: contig of 1421 bp in length  
\* 29759 29858: gap of unknown length  
\* 29859 31038: contig of 1180 bp in length  
\* 31039 31138: gap of unknown length  
\* 31139 32179: contig of 1041 bp in length  
\* 32180 32279: gap of unknown length  
\* 32280 33796: contig of 1517 bp in length  
\* 33797 33896: gap of unknown length  
\* 33897 35380: contig of 1484 bp in length  
\* 35381 35480: gap of unknown length  
\* 35481 37032: contig of 1552 bp in length  
\* 37033 37132: gap of unknown length  
\* 37133 38414: contig of 1282 bp in length  
\* 38415 38514: gap of unknown length  
\* 38515 39948: contig of 1434 bp in length  
\* 39949 40048: gap of unknown length



[illegible]

Db 15439 TTACACACACGGCTGTGGTCTGAACCTGTGAGCCTTGACCTGCCTTCTCCCACTTGTTC

```

Qy 57 SerTyrAlaLysAsnAsnThrAlaAspTrpIleGlnLeuGluLysIleArgAlaProSer 76
Db 15382 -----AGAGCCCTTCC 15371

Qy 77 AsnValSerThrValIleHisIleLeuTyrLeuProGluGluAlaLysGlyGluSerVal 96
Db 15370 AATGTGACGACAGTCATCCATCTCTACCTCCCGAGGAAGCAAGGGAGAGCGTG 15311

Qy 97 GlnPheGlnTrpLysGlnAspSerLeuArgValGlyGluValTyrGluAlaCysTrpAla 116
Db 15310 CAGTTCAGTGGAAACAGGACGCTCGAGTGGGTGAGTGTATGAGGCTGCTGGGCC 15251

Qy 117 Leu 117
Db 15250 CTG 15248

RESULT 11
AC023062 183156 bp DNA linear HTG 05-MAY-2000
LOCUS Mus musculus chromosome 5 clone RP23-466D2 strain C57BL6/J, WORKING
DEFINITION DRAFT SEQUENCE, 8 unordered pieces.
AC023062
AC023062.2 GI:7709916
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Mus musculus.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 183156)
Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
Bouffard,G.G., Dietrich,N.L., Eagle,W.O., Gupta,J., Ho,S.-L.,
Huang,M.C., Idol,J., Lee-Lin,S.-O., Maduro,Q.L., Maduro,V.B.,
Mastrian,S.D., McCloskey,J.C., Morse,E., Ojodu,M.A., Pearson,R.,
Stantripop,S., Summers,T.J., Thomas,J.W., Thomas,P.J.,
Tiongson,E.E., Touchman,J.W., Tran,J.T., Vogt,J.L., Walker,M.A.,
Wetherby,K.D. and Green,E.D.
NISC Mouse Sequencing Initiative
Unpublished
2 (bases 1 to 183156)
Green,E.D.
Direct Submission
Submitted (08-FEB-2000) NIH Intramural Sequencing Center, 8717
GroveMont Circle, Gaithersburg, MD 20877, USA
On May 5, 2000 this sequence version replaced gi:6939141.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.mouse@nhgri.nih.gov
----- Project Information
Center project name: yi
Center clone name: 466D02
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 175371 bases at least Q40
Consensus quality: 178474 bases at least Q30
Consensus quality: 180133 bases at least Q20
Insert size: 179000; agarose-fp
Insert size: 188000; pulse-field-gel
Insert size: 183156; sum-of-contigs
Quality coverage: 6.96x in Q20 bases; agarose-fp
Quality coverage: 6.63x in Q20 bases; pulse-field-gel
Quality coverage: 6.80x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

```

\* as soon as it is available and the accession number will  
\* be preserved.

1 2374: contig of 2374 bp in length  
\* 2375 4490: gap of unknown length  
\* 2375 4490: contig of 2116 bp in length  
\* 4491 14959: gap of unknown length  
\* 4491 14959: contig of 10469 bp in length  
\* 14960 31072: gap of unknown length  
\* 31073 50119: contig of 16113 bp in length  
\* 31073 50119: gap of unknown length  
\* 50120 78196: contig of 19047 bp in length  
\* 50120 78196: gap of unknown length  
\* 78197 109874: contig of 28077 bp in length  
\* 78197 109874: gap of unknown length  
\* 109875 183156: contig of 31678 bp in length  
\* 109875 183156: gap of unknown length  
\* 109875 183156: contig of 73282 bp in length.

FEATURES  
source  
1. .183156  
/organism="Mus musculus"  
/strain="C57BL6/J"  
/db\_xref="taxon:10090"  
/chromosome="5"  
/clone="RP23-466D2"  
/clone\_lib="RPCI mouse BAC library 23"  
BASE COUNT 53719 a 37080 c 37302 g 55052 t 3 others  
ORIGIN

Alignment Scores:  
Pred. No.: 4.84e-17 Length: 183156  
Score: 240.00 Matches: 50  
Percent Similarity: 66.67% Conservative: 4  
Best Local Similarity: 61.73% Mismatches: 9  
Query Match: 38.40% Indels: 18  
DB: 2 Gaps: 2

US-09-897-438B-2 (1-117) x AC023062 (1-183156)

Qy 38 PheSerIleGlySerGly---SerCysArgPheSerTyrSerAspProSerIleThrVal 56  
Db 173566 TTCACACACGGCTGGTCTGAACTGTGACCTTGACCTGCTTCCCTTGTTC--- 173622

Qy 57 SerTyrAlaLysAsnAsnThrAlaAspTrpIleGlnLeuGluLysIleArgAlaProSer 76  
Db 173623 -----AGAGCCCTTCC 173634

Qy 77 AsnValSerThrValIleHisIleLeuTyrLeuProGluGluAlaLysGlyGluSerVal 96  
Db 173635 AATGTGACGACAGTCATCCATCTCTACCTCCCGAGGAAGCAAGGGAGAGCGTG 173694

Qy 97 GlnPheGlnTrpLysGlnAspSerLeuArgValGlyGluValTyrGluAlaCysTrpAla 116  
Db 173695 CAGTTCAGTGGAAACAGGACGCTCGAGTGGGTGAGTGTATGAGGCTGCTGGGCC 173754

Qy 117 Leu 117  
Db 173755 CTG 173757

RESULT 12  
AC124933/c  
LOCUS  
DEFINITION Rattus norvegicus clone CH230-466N17, \*\*\* SEQUENCING IN PROGRESS  
AC124933  
AC124933.2 GI:21952716  
AC124933  
HTG; HTGS\_PHASE1.  
Norway rat.  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 183641)  
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,

Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Ayele, M., Banks, T.,  
Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,  
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,  
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,  
Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,  
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,  
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
Fallis, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,  
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,  
Hernandez, O., Hodgson, A., Hoques, M., Holloway, C., Hollins, B.,  
Homsl, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,  
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudan, S.,  
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,  
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H.,  
Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,  
Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,  
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G.,  
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,  
Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G.,  
Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I.,  
Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,  
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
Tansay, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
Usmani, K., Vasquez, L., Vera, V., Villalob, D., Vinson, R., Wang, Q.,  
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
Williams, G., Williams, A., Wleczky, R., Wooden, S., Worley, K.,  
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
Weinstock, G. and Gibbs, R.

Direct Submission  
Unpublished  
2 (bases 1 to 183641)  
Worley, K.C.

Direct Submission  
Submitted (20-JUN-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 183641)  
Worley, K.C.

Direct Submission  
Submitted (31-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 24, 2002 this sequence version replaced gi:21490053.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: KAVI  
Center clone name: CH230-466N17  
----- Summary Statistics  
Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 111724 bases at least Q30  
Consensus quality: 120970 bases at least Q40  
Consensus quality: 128458 bases at least Q20  
-----

\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 88 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 1107: contig of 1107 bp in length  
\* 1108: gap of unknown length  
\* 1207: contig of 1253 bp in length  
\* 1208: gap of unknown length  
\* 2461: contig of 1444 bp in length  
\* 2561: gap of unknown length  
\* 4004: contig of 1278 bp in length  
\* 4104: gap of unknown length  
\* 5382: contig of 1392 bp in length  
\* 5383: gap of unknown length  
\* 5483: contig of 1392 bp in length  
\* 5484: gap of unknown length  
\* 6874: gap of unknown length  
\* 6875: contig of 1097 bp in length  
\* 8071: gap of unknown length  
\* 8072: contig of 1307 bp in length  
\* 8171: gap of unknown length  
\* 8172: gap of unknown length  
\* 9479: gap of unknown length  
\* 9479: gap of unknown length  
\* 9578: gap of unknown length  
\* 11061: contig of 1483 bp in length  
\* 11062: gap of unknown length  
\* 11161: contig of 1581 bp in length  
\* 11162: gap of unknown length  
\* 12742: contig of 1662 bp in length  
\* 12743: gap of unknown length  
\* 12843: contig of 1662 bp in length  
\* 12843: gap of unknown length  
\* 14505: gap of unknown length  
\* 14505: gap of unknown length  
\* 15734: contig of 1130 bp in length  
\* 15735: gap of unknown length  
\* 15834: contig of 1124 bp in length  
\* 16958: gap of unknown length  
\* 17059: contig of 1646 bp in length  
\* 17059: gap of unknown length  
\* 18705: gap of unknown length  
\* 18705: gap of unknown length  
\* 18805: gap of unknown length  
\* 20420: contig of 1616 bp in length  
\* 20421: gap of unknown length  
\* 21648: contig of 1128 bp in length  
\* 21649: gap of unknown length  
\* 23177: contig of 1429 bp in length  
\* 23178: gap of unknown length  
\* 23278: contig of 1483 bp in length  
\* 24760: gap of unknown length  
\* 24761: gap of unknown length  
\* 24861: contig of 1899 bp in length  
\* 26759: gap of unknown length  
\* 26760: contig of 1440 bp in length  
\* 26859: gap of unknown length  
\* 28299: gap of unknown length  
\* 28300: gap of unknown length  
\* 28400: gap of unknown length  
\* 29592: contig of 1193 bp in length  
\* 29593: gap of unknown length  
\* 30870: contig of 1178 bp in length  
\* 30871: gap of unknown length  
\* 32642: contig of 1672 bp in length  
\* 32643: gap of unknown length  
\* 32742: gap of unknown length  
\* 32743: gap of unknown length  
\* 34387: contig of 1578 bp in length  
\* 34486: gap of unknown length  
\* 36065: contig of 1234 bp in length  
\* 36165: gap of unknown length  
\* 37398: gap of unknown length  
\* 37399: gap of unknown length  
\* 37499: gap of unknown length  
\* 39608: contig of 2110 bp in length  
\* 39609: gap of unknown length  
\* 39709: contig of 1772 bp in length  
\* 41481: gap of unknown length  
\* 41581: contig of 1179 bp in length  
\* 42759: gap of unknown length  
\* 42760: gap of unknown length  
\* 42859: contig of 1852 bp in length  
\* 44711: gap of unknown length  
\* 44811: contig of 1397 bp in length  
\* 46208: gap of unknown length  
\* 46209: gap of unknown length  
\* 46309: contig of 1459 bp in length  
\* 47668: gap of unknown length  
\* 47669: gap of unknown length  
\* 49227: contig of 1360 bp in length  
\* 49228: gap of unknown length  
\* 49328: contig of 1470 bp in length  
\* 50797: gap of unknown length  
\* 50798: gap of unknown length



Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: GDRS

Center clone name: CH230-10G13

----- Summary Statistics

Sequencing vector: plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 70934 bases at least Q40

Consensus quality: 75985 bases at least Q30

Consensus quality: 80731 bases at least Q20

-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 59 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1  
1382: contig of 1382 bp in length  
1383  
1482: gap of unknown length  
1483  
3046: contig of 1564 bp in length  
3047  
3146: gap of unknown length  
3147  
4489: contig of 1343 bp in length  
4490  
4589: gap of unknown length  
4590  
5786: contig of 1197 bp in length  
5787  
5886: gap of unknown length  
5887  
7115: contig of 1229 bp in length  
7116  
7215: gap of unknown length  
7216  
8771: contig of 1556 bp in length  
8772  
10260: contig of 1389 bp in length  
10261  
10360: gap of unknown length  
10361  
11370: contig of 1010 bp in length  
11371  
11470: gap of unknown length  
11471  
12710: contig of 1240 bp in length  
12711  
12810: gap of unknown length  
12811  
14180: contig of 1370 bp in length  
14181  
14280: gap of unknown length  
14281  
16048: contig of 1768 bp in length  
16049  
16148: gap of unknown length  
16149  
17569: contig of 1421 bp in length  
17570  
17670: gap of unknown length  
17671  
18920: contig of 1151 bp in length  
18921  
20209: gap of unknown length  
20210  
20309: contig of 1289 bp in length  
20310  
21633: contig of 1324 bp in length  
21634  
21733: gap of unknown length  
21734  
23269: contig of 1536 bp in length  
23270  
23369: gap of unknown length  
23370  
24647: contig of 1278 bp in length  
24648  
24747: gap of unknown length  
24748  
25800: contig of 1053 bp in length  
25801  
27017: contig of 1117 bp in length  
27018  
27117: gap of unknown length  
27118  
29183: contig of 2066 bp in length  
29184  
30315: gap of unknown length  
30316  
30415: contig of 1032 bp in length  
30416  
32086: contig of 1671 bp in length  
32087  
32186: gap of unknown length  
32187  
33995: contig of 1209 bp in length  
33996  
35663: contig of 2168 bp in length  
35664  
35763: gap of unknown length

35764  
38044: contig of 2281 bp in length  
38045  
38144: gap of unknown length  
38045  
39845: contig of 1701 bp in length  
39846  
39945: gap of unknown length  
39946  
41982: contig of 2037 bp in length  
41983  
42082: gap of unknown length  
42083  
43743: contig of 1661 bp in length  
43744  
43843: gap of unknown length  
43844  
45904: contig of 2061 bp in length  
45905  
46004: gap of unknown length  
46005  
47667: contig of 1663 bp in length  
47668  
47767: gap of unknown length  
47768  
49833: contig of 2066 bp in length  
49834  
49933: gap of unknown length  
49934  
51434: contig of 1501 bp in length  
51435  
51534: gap of unknown length  
51535  
53778: contig of 2244 bp in length  
53779  
53878: gap of unknown length  
53879  
55552: contig of 1674 bp in length  
55553  
55652: gap of unknown length  
55653  
57670: contig of 2018 bp in length  
57671  
57770: gap of unknown length  
57771  
59630: contig of 1860 bp in length  
59631  
59730: gap of unknown length  
59731  
61147: contig of 1417 bp in length  
61148  
61247: gap of unknown length  
61248  
63995: contig of 2748 bp in length  
63996  
64095: gap of unknown length  
64096  
66294: contig of 2199 bp in length  
66295  
66394: gap of unknown length  
66395  
68794: contig of 2400 bp in length  
68795  
68894: gap of unknown length  
68895  
71333: contig of 2639 bp in length  
71334  
71633: gap of unknown length  
71634  
73469: contig of 1836 bp in length  
73470  
73569: gap of unknown length  
73570  
77054: contig of 3485 bp in length  
77055  
77154: gap of unknown length  
77155  
79005: contig of 1851 bp in length  
79006  
79105: gap of unknown length  
79106  
81294: contig of 2189 bp in length  
81295  
81394: gap of unknown length  
81395  
83864: contig of 2470 bp in length  
83865  
83964: gap of unknown length  
83965  
86039: contig of 2075 bp in length  
86040  
86139: gap of unknown length  
86140  
88646: contig of 2507 bp in length  
88647  
88746: gap of unknown length  
88747  
91823: contig of 3077 bp in length  
91824  
91923: gap of unknown length  
91924  
94369: contig of 2446 bp in length  
94370  
94469: gap of unknown length  
94470  
96886: contig of 2417 bp in length  
96887  
96986: gap of unknown length  
96987  
100679: contig of 3693 bp in length  
100680  
100779: gap of unknown length  
100780  
104171: contig of 3392 bp in length  
104172  
104271: gap of unknown length  
104272  
106760: contig of 2489 bp in length

#### Alignment Scores:

Pred. No.:	3 86e-16	Length:	126130
Score:	231.00	Matches:	42
Percent Similarity:	100.00%	Conservative:	3
Best Local Similarity:	93.33%	Mismatches:	0
Query Match:	36.96%	Indels:	0
DB:	2	Gaps:	0

US-09-897-438b-2 (1-117) x AC095877 (1-126130)

QY 73 ArgAlaProSerAsnValSerThrValIleHisIleLeuTyrLeuProGluAlaLys 92

|||||.....:|||||

DB 92827 AGAGCCCCCTCCAACTGAGCACCATCCACATCCTCTACCTTCTCTGAGGACGCCAAA 92886



```

repeat_region 37975..37994
/rpt_family="L1"
repeat_region 40278..40311
/rpt_family="L1"
repeat_region complement(41921..42075)
/rpt_family="ALU"
repeat_region complement(42447..42734)
/rpt_family="ALU"
exon 43126..43185
/notes="GRAIL prediction, score = 95"
/evidence=not_experimental
repeat_region complement(43411..43680)
/rpt_family="ALU"
repeat_region 45704..45991
/rpt_family="ALU"
misc_feature 46064..46434
/notes="match to Human cDNA H60097, 5' end (NID:g1012929)"
repeat_region complement(46604..46897)
/rpt_family="ALU"
misc_feature complement(47023..47412)
/genes="WUGSC:H_RG249A12.1"
repeat_region complement(47023..47412)
/notes="match to Human cDNA, 3' end H59444 (NID:g1012276)"
58125..58163
/rpt_family="L1"
repeat_region 58604..58758
/rpt_family="L1"
repeat_region 59743..59765
/rpt_family="L1"
repeat_region 60210..60516
/rpt_family="ALU"
exon complement(63051..63146)
/genes="WUGSC:H_RG249A12.1"
/notes="GRAIL prediction, score = 100"
/evidence=not_experimental
repeat_region 64242..64284
/rpt_family="L1"
repeat_region 64445..64496
/rpt_family="L1"
repeat_region complement(67753..68011)
/rpt_family="L1"
repeat_region 67986..68889
/rpt_family="L1"
exon complement(70725..70841)
/genes="WUGSC:H_RG249A12.1"
/notes="GRAIL prediction, score = 82"
/evidence=not_experimental
repeat_region 71911..71947
/rpt_family="L1"
repeat_region complement(73631..73919)
/rpt_family="ALU"
repeat_region 75073..75092
/rpt_family="L1"
repeat_region 76743..82366
/rpt_family="L1"
exon 78326..79402
/notes="GRAIL prediction, score = 83"
/evidence=not_experimental
exon 80586..81548
/notes="GRAIL prediction, score = 82"
/evidence=not_experimental
repeat_region complement(80962..81380)
/rpt_family="L1"
repeat_region complement(82581..82608)
/rpt_family="L1"
misc_feature 83877..84447

```

```

Alignment Scores:
Pred. No.: 1,91e-15 Length: 93163
Score: 224.00 Matches: 40
Percent Similarity: 100.00% Conservative: 5
Best Local Similarity: 88.89% Mismatches: 0
Query Watch: 35.84% Indels: 0
DB: 9 Gaps: 0

```

```

US-09-897-438b-2 (1-117) x HSAC000121 (1-93163)

QY 73 ArgAlaProSerAsnValSerThrValIleHisIleLeuTyLeuProGluGluAlaLys 92
|||||
Db 33034 AGAGCCCTTCCAAATGTCAGCACAAATCATCATCTACCTCTCTACCTGAGGAGCCAAA 32975
|||||
QY 93 GlycSerValGlnPheGlnTrpLysGlnAspSerLeuArgValGlyGluValTyGlu 112
|||||
Db 32974 GGGGAGAAATGCCAAATTCAGTGGAGCAAGAAATCTTCGTGTAGGTGAAGTGTATGAA 32915
|||||
QY 113 AlaCystTrpAlaLeu 117
|||||
Db 32914 GCCTGCTGGGCCTTA 32900
|||||

RESULT 15
AC041023/c
LOCUS
DEFINITION Homo sapiens clone RP11-356G3, WORKING DRAFT SEQUENCE, 20 unordered
pieces.
AC041023
VERSION AC041023.2 GI:8016758
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 185996)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Brown,A., Burkett,G.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,W.
Direct Submission
Submitted (11-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 22, 2000 this sequence version replaced gi:7534225.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9601
Center clone name: 356.G.3
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 175155 bases at least Q40

```

Consensus quality: 180623 bases at least Q30  
 Consensus quality: 182712 bases at least Q20  
 Insert size: 188000; agarose-fp  
 Quality coverage: 4.5 in Q20 bases; agarose-fp  
 Quality coverage: 4.6 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 20 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 2941: contig of 2941 bp in length  
 2942 3041: gap of 100 bp  
 3042 5193: contig of 2152 bp in length  
 5194 5293: gap of 100 bp  
 5294 7050: contig of 1757 bp in length  
 7051 7150: gap of 100 bp  
 7151 10452: contig of 3302 bp in length  
 10453 10552: gap of 100 bp  
 10553 13859: contig of 3307 bp in length  
 13860 13959: gap of 100 bp  
 13960 17254: contig of 3295 bp in length  
 17255 17354: gap of 100 bp  
 17355 20523: contig of 3169 bp in length  
 20524 20623: gap of 100 bp  
 20624 25148: contig of 4525 bp in length  
 25149 25248: gap of 100 bp  
 25249 29676: contig of 4428 bp in length  
 29677 29776: gap of 100 bp  
 29777 34077: contig of 4301 bp in length  
 34078 34177: gap of 100 bp  
 34178 39009: contig of 4832 bp in length  
 39010 39109: gap of 100 bp  
 39110 45564: contig of 6455 bp in length  
 45565 45664: gap of 100 bp  
 45665 54035: contig of 8371 bp in length  
 54036 54135: gap of 100 bp  
 54136 63394: contig of 9259 bp in length  
 63395 63494: gap of 100 bp  
 63495 74946: contig of 11452 bp in length  
 74947 75046: gap of 100 bp  
 75047 88353: contig of 13307 bp in length  
 88354 88453: gap of 100 bp  
 88454 104853: contig of 16400 bp in length  
 104854 104953: gap of 100 bp  
 104954 123339: contig of 18386 bp in length  
 123340 123439: gap of 100 bp  
 123440 154730: contig of 31291 bp in length  
 154731 154830: gap of 100 bp  
 154831 185996: contig of 31166 bp in length.

FEATURES

source  
 1. .185996  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="RPCI-11 Human Male BAC"  
 misc\_feature  
 1. .2941  
 /note="assembly\_fragment"  
 misc\_feature  
 3042. .5193  
 /note="assembly\_fragment"  
 misc\_feature  
 5294. .7050  
 /note="assembly\_fragment"  
 misc\_feature  
 7151. .10452  
 /note="assembly\_fragment"  
 misc\_feature  
 10553. .13859  
 /note="assembly\_fragment"  
 misc\_feature  
 13960. .17254  
 /note="assembly\_fragment"  
 misc\_feature  
 17355. .20523

misc\_feature  
 20624. .25148  
 /note="assembly\_fragment"  
 misc\_feature  
 25249. .29676  
 /note="assembly\_fragment"  
 misc\_feature  
 29777. .34077  
 /note="assembly\_fragment"  
 misc\_feature  
 34178. .39009  
 /note="assembly\_fragment"  
 misc\_feature  
 39110. .45564  
 /note="assembly\_fragment"  
 misc\_feature  
 45665. .54035  
 /note="assembly\_fragment"  
 misc\_feature  
 54136. .63394  
 /note="assembly\_fragment"  
 misc\_feature  
 63495. .74946  
 /note="assembly\_fragment"  
 misc\_feature  
 75047. .88353  
 /note="assembly\_fragment"  
 misc\_feature  
 88454. .104853  
 /note="assembly\_fragment"  
 misc\_feature  
 104954. .123339  
 /note="assembly\_fragment"  
 misc\_feature  
 123440. .154730  
 /note="assembly\_fragment"  
 misc\_feature  
 154831. .185996  
 /note="assembly\_fragment"  
 misc\_feature  
 185997. .20523  
 /note="assembly\_fragment"  
 misc\_feature  
 20524. .25148  
 /note="assembly\_fragment"  
 misc\_feature  
 25149. .29676  
 /note="assembly\_fragment"  
 misc\_feature  
 29677. .34077  
 /note="assembly\_fragment"  
 misc\_feature  
 34078. .39009  
 /note="assembly\_fragment"  
 misc\_feature  
 39010. .45564  
 /note="assembly\_fragment"  
 misc\_feature  
 45565. .54035  
 /note="assembly\_fragment"  
 misc\_feature  
 54036. .63394  
 /note="assembly\_fragment"  
 misc\_feature  
 63395. .74946  
 /note="assembly\_fragment"  
 misc\_feature  
 74947. .88353  
 /note="assembly\_fragment"  
 misc\_feature  
 88354. .104853  
 /note="assembly\_fragment"  
 misc\_feature  
 104854. .104953  
 /note="assembly\_fragment"  
 misc\_feature  
 104954. .123339  
 /note="assembly\_fragment"  
 misc\_feature  
 123340. .123439  
 /note="assembly\_fragment"  
 misc\_feature  
 123440. .154730  
 /note="assembly\_fragment"  
 misc\_feature  
 154731. .154830  
 /note="assembly\_fragment"  
 misc\_feature  
 154831. .185996  
 /note="assembly\_fragment"  
 misc\_feature  
 185997. .20523  
 /note="assembly\_fragment"

BASE COUNT 60145 a 34255 c 33646 g 56049 t 1901 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 3.18e-05 Length: 185996  
 Score: 144.00 Matches: 36  
 Percent Similarity: 47.17% Conservative: 14  
 Best Local Similarity: 33.96% Mismatches: 24  
 Query Match: 23.04% Indels: 32  
 DB: 2 Gaps: 4

US-09-897-438b-2 (1-117) x AC041023 (1-185996)

QY 1 GluGlnCysGlyThrIleMetHisGlyAsnAlaValThrPheCysGluProTyrGlyPro 20  
 |||||  
 Db 166372 GACAGTGTGGCGGATATGCGCAATGCCGTACCTCTGTGAACCATATGGCCCA 166313  
 QY 21 ArgGluLeuThrThrThrCysLeuAsnThrThrThrAlaSerValLeuGlnPheSerile 40  
 |||||  
 Db 166312 CGAGAACTGCTAAGTATGCTGCTATTCATGTACA-----TTATTCATTATAGTATT 166262  
 QY 41 GlySerGlySerCysArgPheSerTyr-----SerAspProSerile 54  
 |||||  
 Db 166261 TACTGAGCTTTTGGCCCAACAAAGTTTCTTATGCTCTTAAAGTTCAGATGAAAT--- 166205  
 QY 55 ThrValSerTyrAlaLysAsnAsnThrAlaAspTrpIleGlnLeuGluLysIleArgAla 74  
 |||||  
 Db 166205 ----- 166205  
 QY 75 ProSerAsnValSerThrValIleHisIleLeuTyrLeuProGluAlaLysGlyGlu 94  
 |||||  
 Db 166204 ---TCAATATATTTATGAGTCTCTATTACAAATATCTT---AAATTTGGAGGGGTGAT 166151  
 QY 95 ServGlnPheGlnTrp 100  
 |||||  
 Db 166150 TCTATGCCCTATTTTGG 166133

Search completed: November 6, 2002, 20:18:34  
 Job time : 2907 secs



GenCore version 5.1.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 6, 2002, 19:24:47 ; Search time 67 Seconds  
(without alignments)  
535.540 Million cell updates/sec

Title: US-09-897-438B-2

Perfect score: 625

Sequence: 1 EQCGTIMGHNAVTFCEPYCP.....FQWKODSLRGEVYACWAL 117

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV-xlp

-MODEL-frame+p2n.model -DEV-xlp  
-Q-/cgn2\_1/USPTO.spool/US09897438/runat\_06112002\_101925\_11313/app\_query.fasta\_1.263  
-DB-Issued\_Patents\_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOFC=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS-human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09897438 @CGN\_1\_1\_31 @runat\_06112002\_101925\_11313 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued\_Patents\_NA.\*

1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	625	100.0	11673	4	US-09-334-220-3
2	572	91.5	11580	4	US-09-334-220-4
3	81	13.0	2802	4	US-09-351-814-12
4	78	12.5	414	4	US-08-905-223-151
5	72.5	11.6	3102	4	US-09-336-643A-17
6	72.5	11.6	28473	4	US-08-961-527-83
7	71	11.4	686	2	US-08-627-151A-12
8	68	10.9	331	4	US-09-060-756-308
9	68	10.9	2231	4	US-08-961-527-265
10	67.5	10.8	2937	6	5208144-7
11	67	10.7	741	4	US-09-134-001C-2455
12	66.5	10.6	2909	4	US-08-104-158-1

#### ALIGNMENTS

RESULT 1  
US-09-334-220-3  
; Sequence 3, Application US/09334220  
; Patent No. 6323177  
; GENERAL INFORMATION:  
; APPLICANT: St. Jude's Children's Research Hospital  
; APPLICANT: Curran, Thomas  
; APPLICANT: D'Arcangelo, Gabriella  
; TITLE OF INVENTION: INTERACTION OF REELIN WITH VERY LOW  
; TITLE OF INVENTION: DENSITY LIPOPROTEIN (VLDL) RECEPTOR FOR SCREENING AND  
; TITLE OF INVENTION: THERAPIES  
; FILE REFERENCE: 2427/0F704  
; CURRENT APPLICATION NUMBER: US/09/334,220  
; CURRENT FILING DATE: 1999-06-16  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 3  
; LENGTH: 11673  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-09-334-220-3

Alignment Scores:  
Pred. No.: 4.64e-76 Length: 11673  
Score: 625.00 Matches: 117  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 4 Gaps: 0

US-09-897-438B-2 (1-117) x US-09-334-220-3 (1-11673)

QY 1 GluGlnCysGlyThrIleMetHisGlyAsnAlaValThrPheCysGluProTyrGlyPro 20  
|||||  
Db 970 GAGCAGTGTGGCCACCATCATGCATGCGATGCTACCTTCTGTGAGCGCGCT 1029

c	13	66	10.6	318	4	US-09-240-274-214	Sequence 214, App
	14	66	10.6	1509	2	US-08-791-347-2	Sequence 2, Appli
	15	66	10.6	1985	2	US-08-791-347-1	Sequence 1, Appli
	16	66	10.6	3294	1	US-08-409-995-1	Sequence 1, Appli
	17	66	10.6	3294	3	US-08-685-467-1	Sequence 1, Appli
	18	66	10.6	3294	4	US-08-913-942-1	Sequence 1, Appli
	19	66	10.6	3294	4	US-09-268-347-43	Sequence 43, Appli
	20	66	10.6	5502	3	US-08-836-134-1	Sequence 1, Appli
	21	66	10.6	5502	4	US-09-493-784-1	Sequence 1, Appli
	22	65.5	10.5	7620	1	US-07-767-135-1	Sequence 1, Appli
	23	65.5	10.5	7620	1	US-07-841-652-1	Sequence 1, Appli
c	24	65	10.4	568	2	US-08-657-866-9	Sequence 9, Appli
c	25	65	10.4	568	4	US-09-163-951-9	Sequence 9, Appli
c	26	65	10.4	664	4	US-09-221-017B-469	Sequence 469, App
	27	65	10.4	5679	1	US-08-201-697-1	Sequence 1, Appli
	28	64.5	10.3	697	2	US-08-716-284-1	Sequence 1, Appli
c	29	64.5	10.3	1490	4	US-09-269-040-5	Sequence 5, Appli
c	30	64	10.2	321	4	US-09-240-274-103	Sequence 103, App
c	31	64	10.2	1000	1	US-08-599-252-96	Sequence 96, Appl
c	32	64	10.2	1000	5	PCT-US96-06352-96	Sequence 96, Appl
c	33	64	10.2	1000	5	PCT-US96-06583-96	Sequence 96, Appl
c	34	64	10.2	2025	4	US-09-247-155-58	Sequence 58, Appl
	35	64	10.2	2774	3	US-08-732-433-2	Sequence 2, Appli
	36	63.5	10.2	670	4	US-09-040-984-7	Sequence 7, Appli
	37	63.5	10.2	670	4	US-09-123-912-7	Sequence 7, Appli
c	38	63.5	10.2	670	4	US-09-643-597-7	Sequence 7, Appli
	39	63.5	10.2	670	4	US-09-221-017B-574	Sequence 574, App
	40	63.5	10.2	711	3	US-08-622-046B-2	Sequence 2, Appli
	41	63.5	10.2	711	3	US-08-622-046B-13	Sequence 13, Appl
	42	63.5	10.2	711	4	US-09-100-264-2	Sequence 2, Appli
	43	63.5	10.2	760	3	US-08-768-859A-7	Sequence 7, Appli
	44	63.5	10.2	760	3	US-08-767-820A-7	Sequence 7, Appli
	45	63.5	10.2	760	5	PCT-US95-06157-7	Sequence 7, Appli

```
QY 21 ArgGluLeuThrThrThrCysLeuAsnThrThrThrAlaSerValLeuGlnPheSerIle 40
|||||.....|
Db 1030 CGAGAGCTGACCACCATGCTGAAACACAAACAGCATCTGCTCCAGTTTCCATT 1089

QY 41 GlySerGlySerCysArgPheSerTyrSerAspProSerIleThrValSerTyrAlaLys 60
|||||.....|
Db 1090 GGGTCAGGATGATCGATTAGTTACTCTGACCCAGCATCAGTGTCTCATAGCCAAAG 1149

QY 61 AsnAsnThrAlaAspTrpIleGlnLeuGluLysIleArgAlaProSerAsnValSerThr 80
|||||.....|
Db 1150 AACAAATACCCCTGATGGATTCAGTGGAGAAAATAGAGCCCTTCCCAATGTGAGCACA 1209

QY 81 ValIleHisIleLeuTyrLeuProGluGluAlaLysGlyGluSerValGlnPheGlnTrp 100
|||||.....|
Db 1210 GTCATCCACATCTGTACCTCCCGAGGAAGCCAAAGGGAGAGCGTGCAGTTCCAGTGG 1269

QY 101 LysGlnAspSerLeuArgValGlyGluValTyrGluAlaCysTrpAlaLeu 117
|||||.....|
Db 1270 AAACAGACAGCCTGCGAGTGGGTGAGGTGTATGAGGCCCTGCTGGGCCCTG 1320

RESULT 2
US-09-334-220-4
; Sequence 4, Application US/09334220
; Patent No. 6323177
; GENERAL INFORMATION:
; APPLICANT: St. Jude's Children's Research Hospital
; APPLICANT: Curran, Thomas
; APPLICANT: D'Arcangelo, Gabriella
; TITLE OF INVENTION: INTERACTION OF REELIN WITH VERY LOW
; TITLE OF INVENTION: DENSITY LIPOPROTEIN (VLDL) RECEPTOR FOR SCREENING AND
; FILE REFERENCE: 2427/0F704
; CURRENT FILING DATE: US/09/334,220
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 11580
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-334-220-4

Alignment Scores:
Pred. No.: 1,15e-68 Length: 11580
Score: 572.00 Matches: 106
Percent Similarity: 95.73% Conservative: 6
Best Local Similarity: 90.60% Mismatches: 5
Query Match: 91.52% Indels: 0
DB: 4 Gaps: 0

US-09-897-438B-2 (1-117) x US-09-334-220-4 (1-11580)
QY 1 GluGlnCysGlyThrThrIleMetHisGlyAsnAlaValThrPheCysGluProTyrGlyPro 20
|||||.....|
Db 860 GAACAGTGTGGCCGATTAATGCAATGCAATGCGCTCACCTCTGTGAACCATATGCCCCA 919

QY 21 ArgGluLeuThrThrThrCysLeuAsnThrThrThrAlaSerValLeuGlnPheSerIle 40
|||||.....|
Db 920 CGAAGACTGATTACCAAGCCCTTAATACAAACAGCTTCTCTCTCCCAATTTCCATT 979

QY 41 GlySerGlySerCysArgPheSerTyrSerAspProSerIleThrValSerTyrAlaLys 60
|||||.....|
Db 980 GGGTCAGGTTCAATGTCGCTTTAGTTATTCAGACCCAGCATCATCGTGTATATGCCAAG 1039

QY 61 AsnAsnThrAlaAspTrpIleGlnLeuGluLysIleArgAlaProSerAsnValSerThr 80
|||||.....|
Db 1040 AATAACTCTCGGACATGGATTACGTTAGCTAGAGAAAATAGAGCCCTTCCCAATGTGAGCACA 1099

QY 81 ValIleHisIleLeuTyrLeuProGluGluAlaLysGlyGluSerValGlnPheGlnTrp 100
|||||.....|
Db 1100 ATCATCCATATCTCTACCTTCTCTGAGGAGCCCAAGGGAGAGATGTCCAATTTCAAGTGG 1159

QY 101 LysGlnAspSerLeuArgValGlyGluValTyrGluAlaCysTrpAlaLeu 117
|||||.....|
```

```
Db 1160 AAGCAGAAAATCTTCGTGTAGTGAAGTGTATGAAGCCTGCTGGGCCCTTA 1210

RESULT 3
US-09-351-814-12
; Sequence 12, Application US/09351814
; Patent No. 6352851
; GENERAL INFORMATION:
; APPLICANT: Nielsen, Bjarne Roenfeldt
; APPLICANT: Svendsen, Allan
; APPLICANT: Pedersen, Henrik
; APPLICANT: Vind, Jesper
; APPLICANT: Hendriksen, Hanne Vang
; APPLICANT: Frandsen, Torben Peter
; TITLE OF INVENTION: Glucoamylase Variants
; FILE REFERENCE: 5636.200-US
; CURRENT APPLICATION NUMBER: US/09/351,814
; CURRENT FILING DATE: 1999-07-12
; EARLIER APPLICATION NUMBER: PA 1998 00937
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: PA 1998 01667
; EARLIER FILING DATE: 1998-12-17
; EARLIER APPLICATION NUMBER: 60/093,528
; EARLIER FILING DATE: 1998-07-21
; EARLIER APPLICATION NUMBER: 60/115,545
; EARLIER FILING DATE: 1999-01-12
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 12
; LENGTH: 2602
; TYPE: DNA
; ORGANISM: ASPERGILLUS NIGER
US-09-351-814-12

Alignment Scores:
Pred. No.: 0.41 Length: 2602
Score: 81.00 Matches: 34
Percent Similarity: 36.30% Conservative: 15
Best Local Similarity: 25.19% Mismatches: 54
Query Match: 12.96% Indels: 32
DB: 4 Gaps: 4

US-09-897-438B-2 (1-117) x US-09-351-814-12 (1-2602)
QY 9 GlyAsnAlaValThrPheCysGluProTyrGlyProArgGluLeuThrThrThrCysLeu 28
|||||.....|
Db 2007 GCGCGCACCATCTACGAGCGGTACCCCTACTGGATCCGGCAGCGTGACCTCGACCACGCAAG 2066

QY 29 AsnThrThrThrAlaSerValLeuGlnPheSerIleGlySerGlySerCysArgPheSer 48
|||||.....|
Db 2067 ACCACCGCAGCTGCTAGCAAGACCAACAGCAGTACGTCATCAACCTCTGTACCCTCCC 2126

QY 49 TyrSer-----AspProSerIleThrValSerTyrAlaLysAsn----- 61
::: ||| ::| ::| ::| ::| ::|
Db 2127 ACCGCGCTGGCTGTGACTTTTCGATCTGACAGCTTACCACCACTACGCGGAGAACATCTAC 2186

QY 62 -----AsnThrAlaAspTrpIleGlnLeuGlu 70
|||||.....|
Db 2187 CTGGCTCGGATCGATCTCTCAGTGGGTGACTGGGAACACGCGCATAGCTCTGAGT 2246

QY 71 LysIleArgAlaProSerAsnValSerThrValIleHisIleLeuTyrLeuProGluGlu 90
::: ||| ::| ::| ::| ::| ::|
Db 2247 GCTGACAGATACACTTCTGACGACCGCTCTGTGTATGTCTACTGTGACTCTGCGG----- 2300

QY 91 AlaLysGlyGluSerValGlnPhe-----Gln 99
|||||.....|
Db 2301 ---GCTGTTGAGTCGTTTGTAGTACAAGTTTATCCGCAATTGAGACGATGACTCCGTGGAG 2357

QY 100 TrpLysGlnAspSerLeuArgValGlyGluValTyrGluAlaCys 114
|||||.....|
Db 2358 TGGGAGAGTGAATCCCAACCGAGAGATACACCGTTCTCTCAGCGGTGC 2402

RESULT 4
```

US-08-905-223-151  
; Sequence 151, Application US/08905223  
; Patent No. 6222029  
; GENERAL INFORMATION:  
; APPLICANT: Edwards, Jean-Baptiste D.  
; APPLICANT: Duclert, Aymeric  
; APPLICANT: Lacroix, Bruno  
; TITLE OF INVENTION: 5' ESTS FOR SECRETED PROTEINS  
; NUMBER OF SEQUENCES: 503  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobber, Martens, Olson & Bear  
; STREET: 501 West Broadway  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92101-3505  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy Disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: Win95  
; SOFTWARE: Word  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/905,223  
; FILING DATE:  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Israel, Ned A.  
; REGISTRATION NUMBER: 29,655  
; REFERENCE/DOCKET NUMBER:  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 235-8550  
; TELEFAX: (619) 235-0176  
; INFORMATION FOR SEQ ID NO: 151:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 414 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: DOUBLE  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: CDNA  
; ORIGINAL SOURCE:  
; ORGANISM: Homo Sapiens  
; TISSUE TYPE: Brain  
; FEATURE:  
; NAME/KEY: sig\_peptide  
; LOCATION: 262..369  
; IDENTIFICATION METHOD: Von Heijne matrix  
; OTHER INFORMATION: score 3.9  
; OTHER INFORMATION: seq SICLSLLIPGPKP/LV  
US-08-905-223-151

Alignment Scores:  
Pred. No.: 0.0639 Length: 414  
Score: 78.00 Matches: 28  
Percent Similarity: 41.88% Conservative: 21  
Best Local Similarity: 23.93% Mismatches: 48  
Query Match: 12.48% Indels: 20  
DB: 4 Gaps: 4

US-09-897-438B-2 (1-117) x US-08-905-223-151 (1-414)

QY 3 CysGlyThrIleMethHisGlyAsnAlaValThrPheCysGluProTyrGlyProArgGlu 22  
Db 48 TGTGGCTCTACT-----GTTGATGCCCTGCTTCTATTGAAAGGCCACACCGGTAC 101  
QY 23 LeuThrThrThrCysLeuAsnThrThrThrAlaSerValLeuGlnPheSerIleGlySer 42  
Db 102 GTGTCAGCACACAGCGTTCGCGTGAATGAGGATTCCTCTACAGATAGACTTC---GCT 158  
QY 43 GlySerCysArgPheSerTyrSerAspProSerIleThrValSerTyrAlaLysAsnAsn 62  
Db 159 GCTCCCGCTCAGTCACAGACTCTGTATGCGATTAATGGAATACCTCAGTAGATCTT 218  
QY 63 ThrAlaAspTrpIleGlnLeuGluLysIleArgAlaProSerAsnValSer---ThrVal 81

Db 219 GGATTGTCATGGCACCATTGGTAAGGACTGTCTGCTACCAATGTGGAATGAGTCTGC 278  
QY 82 IleHisIle-----Leu 85  
Db 279 TATCATCTGCAAGGATCCCTGGTGTCTACAGACADWTCACAAAGTGGACTAGAACACTCTG 338  
QY 86 TyrLeuProGluGluAlaLysGlyGluSerValGlnPheGlnTrpLysGln 102  
Db 339 CCTCTCCTCTCTTATACCAGGTCCCAAGCCACTCGTTTCGTTGGGCATCAA 389  
RESULT 5  
US-09-336-643A-17  
; Sequence 17, Application US/09336643A  
; Patent No. 6399761  
; GENERAL INFORMATION:  
; APPLICANT: Miller, Andrew P.  
; APPLICANT: Curran, Mark Edward  
; APPLICANT: Hu, Ping  
; APPLICANT: Rutter, Marc  
; APPLICANT: Wang, Jian-Wang  
; TITLE OF INVENTION: No. 6399761 Human Potassium Channels  
; FILE REFERENCE: SEQ-15P  
; CURRENT APPLICATION NUMBER: US/09/336,643A  
; CURRENT FILING DATE: 1999-06-18  
; PRIOR APPLICATION NUMBER: 60/076,687  
; PRIOR FILING DATE: 1998-08-07  
; PRIOR APPLICATION NUMBER: 60/116,448  
; PRIOR FILING DATE: 1999-01-19  
; PRIOR APPLICATION NUMBER: PCT/US99/03826  
; PRIOR FILING DATE: 1999-02-22  
; NUMBER OF SEQ ID NOS: 87  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17  
; LENGTH: 3102  
; TYPE: DNA  
; ORGANISM: H. sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (274)...(1705)  
; OTHER INFORMATION: K+Hnov11  
US-09-336-643A-17

Alignment Scores:  
Pred. No.: 8.26 Length: 3102  
Score: 72.50 Matches: 34  
Percent Similarity: 37.01% Conservative: 13  
Best Local Similarity: 26.77% Mismatches: 56  
Query Match: 11.60% Indels: 25  
DB: 4 Gaps: 5

US-09-897-438B-2 (1-117) x US-09-336-643A-17 (1-3102)

QY 2 GlnCysGlyThrIleMethHisGlyAsnAlaValThrPheCysGluProTyrGlyProArg 21  
Db 336 CAATGTGGCGGCTTCAAGAGAGGCTGCGCTCCACACCGCTGCTGGCT-TCCCCGAGA 394  
QY 22 GluLeuThrThrThrCysLeuAsnThrThrAla-----SerValLeu 36  
Db 395 CGCGCCTGGCGCGCTTGTCTCTGCCACTCGCGGAGGCCATTCTGGAGCTTGGCATG 454  
QY 37 GlnPheSerIleGlySerGlySerCysArgPheSer-----TyrSerAspProSer 53  
Db 455 ACTACGACGACGTCACCGGGAGTTCCTACCGCCCAACCCGAGCTCTTCCCTTACG 514  
QY 54 IleThrValSerTyrAlaLysAsnAsnThrAlaAspTrpIleGlnLeuGluLysIleArg 73  
Db 515 TGCTGCTATTCTATCACACCGGCAAGCTTCACGTCATGG-----CTGAGCTATGCTCT 568  
QY 74 AlaProSer-----AsnValSerThr-ValIleHisIleLeuTyrLe 87  
Db 569 TCCTCTCAGCCAGGAGATCGAGTACTTGGGGCATCAACAGGATCTTCTTATTGACTCTGCT 628

```
Qy 87 uProGluGluAlaLysGlyGluSerValGlnPheGlnTrpLysGlnAspSerLeuArgVa 107
|          :|||  |||  |||  |||
Db 629 G-----CAGCTACAGTACCATTGCGCGCAAAAGTAGAGCCCGAGC 667

Qy 107 lGlyGluValTyrgluAla 113
|          |||
Db 668 AGGAGAAAGTGGGACGACA 686

RESULT 6
US-08-961-527-83/C
; Sequence 83, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 83:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28473 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-527-83

Alignment Scores:
Pred. No.: 249 Length: 28473
Score: 72.50 Matches: 22
Percent Similarity: 41.76% Conservative: 16
Best Local Similarity: 24.18% Mismatches: 33
Query Match: 11.60% Indels: 20
DB: 4 Gaps: 3

US-09-897-438B-2 (1-117) x US-08-961-527-83 (1-28473)
Qy 14 PheCysGluProTyrglyProArgGluLeuThrThrThrCysLeuAsnThrThrAla 33
|          |||  |||  |||  |||
Db 28317 TATTGTCGCGCAACAAGGAAAGAACCTTCGTTTCAGTGCAGCGCAACACAAAGCTGT 28258

Qy 34 SerValLeuGlnPhe-SerIleGlySerGlySerCysArgPheSerTyrrSerAspProSe 53
|          |||  |||  |||  |||
Db 28257 AGAGTTTCAGATATTTGAATCTC-----TTCAATGCTAAAGATGATAA 28216

Qy 53 rIleThrVal---SerTyrrAlaLysAsnThrAlaAspTrpIleGln----- 68
|          |||  |||  |||  |||
Db 28215 GGTAACTGTTCTTTCATTGCAATTCGCGGAAACAGCTGTTTACTTGGAAAAAATGGGTCT 28156

Qy 69 -----LeuGluLysIleArgAlaProSerAsnValSerTh 80
```

```
Db 28155 ACGCGCGCATCAATCTTACTACAAGATTGAACGTATCCGGAATCAACAGTGTCTGTTTA 28096
|          |||  |||  |||  |||
Qy 80 rValIleHisIleLeuTyrrLeuProGluGlu 90
|          |||  |||  |||  |||
Db 28095 TATCTACCATACATCTCTACATCCAGAACAA 28065

RESULT 7
US-08-627-151A-12
; Sequence 12, Application US/08627151A
; Patent No. 5866341
; GENERAL INFORMATION:
; APPLICANT: SPINELIA, Dominic
; APPLICANT: BECHERER, Kathleen
; APPLICANT: BROWN, Steven
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: SCREENING DRUG LIBRARIES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gen-Probe Incorporated
; STREET: 10210 Genetic Center Drive
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/627,151A
; FILING DATE: 03-APR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fisher, Carlos A.
; REGISTRATION NUMBER: 36,510
; REFERENCE/DOCKET NUMBER: CB1016
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-410-8926
; TELEFAX: 619-410-8928
; TELEX:
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 686 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-627-151A-12

Alignment Scores:
Pred. No.: 132 Length: 686
Score: 71.00 Matches: 24
Percent Similarity: 41.49% Conservative: 15
Best Local Similarity: 25.53% Mismatches: 31
Query Match: 11.36% Indels: 24
DB: 2 Gaps: 5

US-09-897-438B-2 (1-117) x US-08-627-151A-12 (1-686)
Qy 26 ThrCysLeuAsnThrThrThrAlaSerValLeuGlnPheSerIleGlySerGlySerCys 45
|          |||  |||  |||  |||
Db 37 AGCTGCTG-----GTCTGCTGCAGGTGGCAAGCTCTGGGAACATG 78

Qy 46 ArgPheSerTyrrSerAspProSerIleThrValSerTyrrAlaLysAsnThrAlaAsp 65
|          |||  |||  |||  |||
Db 79 AAGTCTC---TTGCAGGAGCCACCTCGCTCGCTCCGACTACATGAGCATCTCTACTTGGCAG 135

Qy 66 TrpIleGlnLeuGluLysIleArgAlaProSerAsnValSerThrValIleHisIleLeu 85
|          |||  |||  |||  |||
```

RESULT 9  
US-08-961-527-265/c  
; Sequence 265, Application US/08961527  
; Patent No. 6420135  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch

```

US-09-897-438B-2 (1-117) x US-08-961-527-265 (1-2231)
Qy      6 IleMethHisGlyAsnAlaValThrPheCysGluPro----- 17
      |||   :::   |||   |||
Db      577 ATCTTTTAGTAAGACAGAGTCTTGTTCACGAGCCCTATTCCAAAGTTACATTTGCGCCAT 518
      |||   |||   |||   |||
Qy      18 -----TyrGlyProArgGluLeuThrThrCysLeuAsnThrThrAla--- 33
      |||   |||   |||   |||   |||   |||   |||   |||
Db      517 CTAAATAGCTATGCTAAAAAGAGGCGCAGAGATTCTCTAGGACCAGGATGCCATG 458
      |||   |||   |||   |||   |||   |||   |||   |||
Qy      34 -----SerValLeuGlnPheSerIleGlySerGlySerCysArgPheSerTyrSer 50
      |||   |||   |||   |||   |||   |||   |||   |||
Db      457 TTGGGTCTAAATCATGTCTCTTTGGTCTGGGTCTTTTCCCAACGAATAGTCGCTTC 398
      |||   |||   |||   |||   |||   |||   |||   |||
Qy      51 AspProSerIle-----ThrValSerTyrAlaLysAsnAsnThrAlaAspTrpIle 67
      |||   |||   |||   |||   |||   |||   |||   |||
Db      397 GTTCCCTCTGTTTTAACTCGGACGGTGAGTCAGTCTTTTAAATCCC----- 353
      |||   |||   |||   |||   |||   |||   |||   |||
Qy      68 GlnLeuGluLysIleArgAlaProSerAsnValSerThrValIleHisIle----- 84
      |||   |||   |||   |||   |||   |||   |||   |||
Db      352 ATGCTACGAAGGTT---GTTCCCAAAAGGAAGAGACCTGCTTCACCTGATGAATCAG 296
      |||   |||   |||   |||   |||   |||   |||   |||
Qy      85 -----LeuTyrLeuProGluGluAlaLysGlyGluSerValGln 97
      |||   |||   |||   |||   |||   |||   |||   |||
Db      295 CCTATCTACCAAGTTTGTATGCTGATCACACCAAGAACTGTGAGAAGCGGCAGCTGAT 236
      |||   |||   |||   |||   |||   |||   |||   |||
Qy      98 PheGln 99
      |||   |||   |||   |||   |||   |||   |||   |||
Db      235 TTTCGAA 230

```

RESULT 10  
5208144-7  
; Patent No. 5208144  
; APPLICANT: SMITH, JOHN A.; RAYCHORDHURY, RAKTIMA; NILES, JOHN L.  
; TITLE OF INVENTION: METHOD FOR DETECTION OF HUMAN DNA  
; CONTAINING THE GENE ENCODING LOW DENSITY LIPOPROTEIN RECEPTOR  
; NUMBER OF SEQUENCES: 42  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/396,697  
; FILING DATE: 22-AUG-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 313,682  
; FILING DATE: 22-FEB-1989  
; APPLICATION NUMBER: 235,211  
; FILING DATE: 23-AUG-1988  
; SEQ ID NO: 7:  
; LENGTH: 2937  
5208144-7

Alignment Scores:  
Pred. No.: 37.9 Length: 2937  
Score: 67.50 Matches: 28  
Percent Similarity: 36.50% Conservative: 22  
Best Local Similarity: 20.44% Mismatches: 59  
Query Match: 10.80% Indels: 28  
Gaps: 6  
US-09-897-438B-2 (1-117) x 5208144-7 (1-2937)  
Qy 1 GluGlnCysGlyThrIleMethHisGlyAsnAlaValThrPheCysGluPro---TyrGly 19  
Db 682 CAGAACTGTACCCAGCTGACGAGCGAGCGCTTATCTCTCTCTCGACACCGGGTTCAA 741  
Qy 20 ProArgGluLeu---ThrThrThrCysLeuAsnThrThrAlaSerValLeuGlnPhe 38  
Db 742 CCCAGTACTTTGGACAAAATTCCTGCAAGACATCAATGAATGTGAG-----GAGTTT 795  
Qy 39 SerIleGlySerGlySerCysArgPheSerTyrSerAsp----- 51  
Db 796 GCATCTGCCCGACAGCTGCGAAACAGCAAGAAAGTATGATGTTCTGTGTGTCAT 855  
Qy 52 -----ProSerIleThrValSerTyrAlaLysAsnAsnThrAlaAspTrp----- 66  
Db 856 GCGTTCAAGTCTATCAGTACCATATATGAGAACGGTGTGCGAGCTGATGGAAGCCCTCT 915  
Qy 67 -----IleGlnLeuGluLysIleArgAlaProSerAsnValSer 79  
Db 916 CTCCTGCTCCTCGCGAGAAATGCGAATCCGGAAGTACATACCTCTCGAGAATGTC 975  
Qy 80 ThrValIleHisIleLeuTyrLeuProGluGluAlaLysGlyGluSerValGlnPheGln 99  
Db 976 TCAGAG-----TACTGGAAGAGGAGGAGCATATCCAACTATTGACTATGAC 1023  
Qy 100 TrpLysGlnAspSerLeuArgValGlyGluValTyr-GluAlaCysTrp 115  
Db 1024 TGGATCCCGACACATAGGCTCAGTGTGTCTATTACACTGTGCTGG 1072

RESULT 11  
US-09-134-001C-2455/c  
; Sequence 2455, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 2455  
; LENGTH: 741  
; TYPE: DNA  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-2455  
Alignment Scores:  
Pred. No.: 5.37 Length: 741  
Score: 67.00 Matches: 20  
Percent Similarity: 47.37% Conservative: 7  
Best Local Similarity: 35.09% Mismatches: 16  
Query Match: 10.72% Indels: 14  
Gaps: 3  
US-09-897-438B-2 (1-117) x US-09-134-001C-2455 (1-741)  
Qy 9 GlyAsnAlaValThrPheCysGluProTyrGlyProArgGluLeuThrThrCysLeu 28  
Db 380 GGAGATGCTTTTCTTCTATTGCTAAACCGATAGACCATATCCA----- 339  
Qy 29 AsnThrThrThrAlaSerValLeuGlnPheSerIleGlySerGlySerCysArgPhe--- 47  
Db 338 CATCTACATCCGCGATGGTTTA-----CTTGGCGCTGGAGGATGTTCTTTCAA 288  
Qy 48 -----SerTyrSerAspProSerIleThrValSerTyrAlaLys 60  
Db 287 AAAGTTTAAATAGTAGTCTGATCCAAATCCACATATCTTTGAAAAA 237  
RESULT 12  
US-08-104-158-1/c  
; Sequence 1, Application US/08104158  
; Patent No. 6215042  
; GENERAL INFORMATION:  
; APPLICANT: Willmitzer, Lothar  
; APPLICANT: Sonnewald, Uwe  
; APPLICANT: Kossmann, Jens  
; APPLICANT: Mueller-Roeber, Bernd  
; APPLICANT: Visser, Richard Gerardus Franciscus  
; APPLICANT: Jacobsen, Evert  
; TITLE OF INVENTION: PLASMIDS CONTAINING DNA-SEQUENCES THAT  
; TITLE OF INVENTION: CAUSE CHANGES IN THE CARBOHYDRATE CONCENTRATION AND THE  
; TITLE OF INVENTION: CARBOHYDRATE COMPOSITION IN PLANTS, AS WELL AS PLANT CELLS  
; TITLE OF INVENTION: AND PLANTS CONTAINING THESE PLASMIDS  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen  
; STREET: 1180 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-8403  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/104,158  
; FILING DATE: 13-AUG-1993  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP PCT/EP92/00302  
; FILING DATE: 11-FEB-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE P 41 04 782.6  
; FILING DATE: 13-FEB-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meilman, Edward A.  
; REGISTRATION NUMBER: 24,735  
; REFERENCE/DOCKET NUMBER: FA-1996 PCT (951-91)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-382-0700

```
; TELEFAX: 212-382-0888
; TELEX: 236925
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2909 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Solanum tuberosum cv. Desiree
; STRAIN: Desiree
; DEVELOPMENTAL STAGE: growing tuber
; TISSUE TYPE: tuber
; CELL TYPE: total tuber
; IMMEDIATE SOURCE:
; LIBRARY: cDNA of total tuber mRNA in pUC 19 (Hinc II)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..1699
; OTHER INFORMATION: /note= "for Branching enzyme I
; OTHER INFORMATION: (partial) truncated protein; 97,11 % identity to
; OTHER INFORMATION: active potato branching enzyme"
; US-08-104-158-1

Alignment Scores:
Pred. No.: 51.5 Length: 2909
Score: 66.50 Matches: 28
Percent Similarity: 45.79% Conservatve: 21
Best Local Similarity: 26.17% Mismatches: 37
Query Match: 10.64% Indels: 21
DB: 4 Gaps: 6

US-09-897-438b-2 (1-117) x US-08-104-158-1 (1-2909)
Qy 26 ThrCysLeuAsnThrThrAlaSerVal-----LeuGlnPheSer----- 39
Db 1589 ACATGCCAGATACATCTCTTTGTCATAGGAGAAATGCAATGCTGTGCACCGACAA 1530
Qy 40 ---IleGlySerGlySerCysArgPheSerTyrSerAspProSerIleThrValSerTyr 58
Db 1529 TAGACTGATCATGCTCTCCGCATATGCTATACACT-----TCTGTATATCTCTAT 1476
Qy 59 AlaLysAsnAsn-----ThrAlaAspTrpIleGlnLeuGluLysIleArgAlaProSer 76
Db 1475 TTGTCAAACTCGATGTTTACTTCTTCATGGACCAATCTTCATCATCTTATTTTAAAT 1416
Qy 77 AsnValSerThrValIleHisIle-----LeuTyrLeuPro 88
Db 1415 AATCATCCACTTATCTGGGATTCGCCAGCGGTATCAAAACCAATTCCTCCCT 1356
Qy 89 GluGluAlaLys-----GlyGluSerValGlnPheGlnTrpLysGlnAspSerLeuArg 106
Db 1355 CAGAAACAGGCGCGCTAAGCGCCGATACCAAGAAACATCTTCGCAATAACACTTTCGT 1296
Qy 107 ValGlyGluValTyrGluAla 113
Db 1295 CTGGGAAATCTTGTGAATCA 1275

RESULT 13
US-09-240-274-214/c
; Sequence 214, Application US/09240274
; Patent No. 6255455
; GENERAL INFORMATION:
; APPLICANT: Siegel, Donald L.
; TITLE OF INVENTION: Rh(D)-BINDING PROTEINS AND MAGNETICALLY ACTIVATED CELL
; TITLE OF INVENTION: SORTING METHOD FOR PRODUCTION THEREOF
; FILE REFERENCE: 09596-4202
; CURRENT APPLICATION NUMBER: US/09/240, 274
; CURRENT FILING DATE: 1999-01-29
; EARLIER APPLICATION NUMBER: 60/081,380
```

```
; EARLIER FILING DATE: 1998-04-10
; EARLIER APPLICATION NUMBER: 60/028,550
; EARLIER FILING DATE: 1996-10-11
; NUMBER OF SEQ ID NOS: 224
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 214
; LENGTH: 318
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: anti-Rh(D) antibody clone SH44
; US-09-240-274-214

Alignment Scores:
Pred. No.: 2.02 Length: 318
Score: 66.00 Matches: 25
Percent Similarity: 38.32% Conservatve: 16
Best Local Similarity: 23.36% Mismatches: 48
Query Match: 10.56% Indels: 18
DB: 4 Gaps: 4

US-09-897-438b-2 (1-117) x US-09-240-274-214 (1-318)
Qy 15 CysGluProTyrGlyProArgGluLeuThrThrThrCysLeuAsnThrThrThrAlaSer 34
Db 303 TGTGTCCTCCAGGCGCGAAAGAGAGATTACTTAACCTCTGTGGCAGTAGTAAGTTGGCAA 244
Qy 35 ValLeuGlnPheSerIleGlySerGlySer-----CysArgPheSerTyrSer--- 50
Db 243 ATCTTCAGGTTCGACACTGCTGTGAGAGTGAACCTGTGCCAGATCCACTGGCACT 184
Qy 51 -----AspProSerIleThrValSerTyrAla-----LysAsnAsnThr 63
Db 183 GAACCTTGATGGGACCCCACTTTGTAACCTGGATATATCATCATCAGGAGACAGGGGC 124
Qy 64 AlaAspTrpIleGlnLeuGluLysIleArgAlaProSerAsnValSerThrValIleHis 83
Db 123 TTTTCCAGGCTCTGTTGATACCAATTCAGAACCTGGGAATGCTCTGACTTGC----- 70
Qy 84 IleLeuTyrLeuProGluGluAlaLysGlyGluSerValGlnPheGlnTrpLysGlnAsp 103
Db 69 -----CCGCAAGTGATGATGACTCTCTCTCTACAGATCGACAGGGA 25
Qy 104 SerLeuArgValGlyGluVal 110
Db 24 CGATGGAGACTGGGTGAGCTC 4

RESULT 14
US-08-791-347-2
; Sequence 2, Application US/08791347
; Patent No. 5885776
; GENERAL INFORMATION:
; APPLICANT: Stone, Edwin M.
; APPLICANT: Sheffield, Val C.
; APPLICANT: Alward, Wallace L.M.
; TITLE OF INVENTION: GLAUCOMA COMPOSITIONS AND THERAPEUTIC
; TITLE OF INVENTION: AND DIAGNOSTIC USES THEREFOR
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESS: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/791,347
; FILING DATE: 30-JAN-1997
```

```
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold, Beth E.
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: UIA-010.26
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1509 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-08-791-347-2

Alignment Scores:
Pred. No.: 22.1 Length: 1509
Score: 66.00 Matches: 24
Percent Similarity: 44.86% Conservative: 24
Best Local Similarity: 22.43% Mismatches: 35
Query Match: 10.56% Indels: 24
Gaps: 5

US-09-897-438B-2 (1-117) x US-08-791-347-2 (1-1509)
Qy 15 CysGluProTyrglyPro-----ArgGluLeuThrThrCysLeu--- 28
Db 22 TGCTCCAGCTTTGGGCGCTGAGATGCCAGCTGCCAGCTGCTTGGCGCTGCTGGTG 81
Qy 29 -----AsnThrThrAlaSerValLeuGlnPheSerIleGlySerGlySerCys 45
Db 82 TGGGATGTGGGGCGCAGACAGCTCAGCTCAGGAAGGCCAATGACCAGAGTGGCGGATGC 141
Qy 46 ArgPheSerTygSer-----AspPro----- 52
Db 142 CAGTATACCTTCAGTGTGGCGCAGTCCCAATGAATCCAGCTGCCAGAGCAGAGCGCC 201
Qy 53 ---SerIleThrValSerTyrglyAlaLysAsnThrAlaAspTrpIleGlnLeuGluLys 71
Db 202 ATGCAGTCATCCATACTTACAGAAAGACAGACAGACCCCAACGCTTAGACCTGGAGGCC 261
Qy 72 IleArgAlaPro---SerAsnValSerThrValIleHisIleLeuTygLeuProGluGlu 90
Db 262 ACCAAAGCTCGACTCAGCTCCCTGGAGAGCGCTCCTCCACCAATTTGACCTTGGACCGAGCT 321
Qy 91 AlalysGlyGluSerValGln 97
Db 322 GCCAGGCCCGCAGAGAGACCCAG 342

RESULT 15
US-08-791-347-1
; Sequence 1, Application US/08791347
; Patent No. 5885776
; GENERAL INFORMATION:
; APPLICANT: Stone, Edwin M.
; APPLICANT: Sheffield, Val C.
; APPLICANT: Alward, Wallace L.M.
; TITLE OF INVENTION: GLAUCOMA COMPOSITIONS AND THERAPEUTIC
; TITLE OF INVENTION: AND DIAGNOSTIC USES THEREFOR
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/791,347
; FILING DATE: 30-JAN-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold, Beth E.
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: UIA-010.26
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1985 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-791-347-1

Alignment Scores:
Pred. No.: 33.6 Length: 1985
Score: 66.00 Matches: 24
Percent Similarity: 44.86% Conservative: 24
Best Local Similarity: 22.43% Mismatches: 35
Query Match: 10.56% Indels: 24
Gaps: 5

US-09-897-438B-2 (1-117) x US-08-791-347-1 (1-1985)
Qy 15 CysGluProTyrglyPro-----ArgGluLeuThrThrCysLeu--- 28
Db 89 TGCTCCAGCTTTGGGCGCTGAGATGCCAGCTGCCAGCTGCTTGGCGCTGCTGGTG 148
Qy 29 -----AsnThrThrAlaSerValLeuGlnPheSerIleGlySerGlySerCys 45
Db 149 TGGGATGTGGGGCGCAGACAGCTCAGCTCAGGAAGGCCAATGACCAGAGTGGCGGATGC 208
Qy 46 ArgPheSerTygSer-----AspPro----- 52
Db 209 CAGTATACCTTCAGTGTGGCGCAGTCCCAATGAATCCAGCTGCCAGAGCAGAGCGCC 268
Qy 53 ---SerIleThrValSerTyrglyAlaLysAsnThrAlaAspTrpIleGlnLeuGluLys 71
Db 269 ATGTCAGTCATCCATACTTACAGAAAGACAGACAGACCCCAACGCTTAGACCTGGAGGCC 328
Qy 72 IleArgAlaPro---SerAsnValSerThrValIleHisIleLeuTygLeuProGluGlu 90
Db 329 ACCAAAGCTCGACTCAGCTCCCTGGAGAGCGCTCCTCCACCAATTTGACCTTGGACCGAGCT 388
Qy 91 AlalysGlyGluSerValGln 97
Db 389 GCCAGGCCCGCAGAGAGACCCAG 409

Search completed: November 6, 2002, 20:54:57
Job time : 76 secs
```



GenCore version 5.1.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 6, 2002, 18:12:02 : Search time 299 Seconds  
(without alignments)  
881.217 Million cell updates/sec

Title: US-09-897-438B-2

Perfect score: 625

Sequence: 1 EQCGTIMGNAVTCEPYGP.....FQWKQDSLRVGEVYACWAL 117

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DRV=xlp  
-Q=/cgn2\_1/USP10.spool/US09897438/runat\_06112002\_101923\_11281/app\_query.fasta\_1.263  
-DB=N\_Geneseq\_101002 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS-human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTPMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09897438 @CGL1.1.79\_runat\_06112002\_101923\_11281 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOB=6 -DELEXT=7

Database :

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	625	100.0	351	24	ABL40165
2	625	100.0	2745	24	ABD22754
3	625	100.0	11673	24	ABA92603
4	625	100.0	11673	24	ABN99284
5	572	91.5	11580	24	ABN96939
6	572	91.5	11580	24	ABA92604
7	572	91.5	11632	23	AAS89484
8	499	79.8	2025	24	AAD22779
9	499	79.8	2274	24	AAD22753
10	144	23.0	413	22	AAK07720
11	144	23.0	413	22	AAK33553
12	144	23.0	413	24	ABS08422
13	109.5	17.5	4576	23	AAS89475
14	107	17.1	87	22	AAK20384
15	107	17.1	87	22	AAK46490
16	107	17.1	87	24	ABS20836
17	83	13.3	1174	21	AAF11251
18	83	13.3	1845	5	AAK40165
19	83	13.3	3408	5	AAK40166
20	81.5	13.0	1404	23	AAS89482
21	81	13.0	2258	21	AAF11248
22	81	13.0	2600	9	AAK2019
23	81	13.0	2602	21	AAK87843
24	81	13.0	2602	22	AAC85109
25	80	12.8	9477	23	ABL14686
26	78	12.5	414	20	AAK51900
27	73.5	11.8	700	22	AAK92336
28	73	11.7	1820	11	AAK04731
29	73	11.7	2615	11	AAK04730
30	72.5	11.6	3102	20	AAZ11905
31	72.5	11.6	28473	19	AAV52216
32	72	11.5	795	24	ABK75454
33	72	11.5	1654	21	AAK56613
34	71.5	11.4	932	23	ABV22646
35	71.5	11.4	932	23	ABV28470
36	71.5	11.4	2153	23	AAK73449
37	71	11.4	586	21	AAF11295
38	71	11.4	2562	22	AAC84443
39	70.5	11.3	813	24	ABN70601
40	70.5	11.3	837	24	ABK66616
41	70.5	11.3	871	23	ABV23361
42	70.5	11.3	871	23	ABV29216
43	70.5	11.3	1237	21	AAC51174
44	70.5	11.3	1239	21	AAC47796
45	70	11.2	531	22	AAK13506

ALIGNMENTS

RESULT 1  
ABL40165  
ID ABL40165 standard; DNA; 351 BP.  
XX  
AC ABL40165;  
XX  
DT 21-MAY-2002 (first entry)  
XX  
DE Mouse reelin protein CR-50 epitope region encoding DNA SEQ ID NO:1.  
XX  
KW Mouse; reelin protein CR-50 epitope region; elucidation; neuron;  
XX cerebral disturbance; reelin protein; neuroprotective; gene; ds.  
OS Mus musculus.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..351  
FT /tag= a

FT	/partial
FT	/product= "reelin protein CR-50 epitope region"
FT	/note= "no start or stop codons given"
XX	
PN	JP2002017361-A.
XX	
PD	22-JAN-2002.
XX	
PF	04-JUL-2000; 2000JP-0202801.
XX	
PR	04-JUL-2000; 2000JP-0202801.
XX	
PA	(RIKA ) RIKAGAKU KENKYUSHO.
XX	
DR	WPI; 2002-221707/28.
XX	
DR	P-PSDB; ABB06244.
XX	
PT	Reelin protein CR-50 epitope region, useful for diagnosis and treatment
PT	of cerebral disturbance -
XX	
PS	Claim 5; Page 11; 16pp; Japanese.
XX	
CC	The present sequence encodes the mouse reelin protein CR-50 epitope
CC	region, which contains the CR-50 antibody recognition site and is free
CC	from F-spondin domains and repetitive sites. Also described are: (1) an
CC	expression vector comprising a polynucleotide encoding a reelin protein;
CC	epitope region; (2) host cells with transfected the expression vector;
CC	(3) polypeptides prepared by culture of the host cells; and (4)
CC	polynucleotides comprising the 351 base sequence given in ABL40165 which
CC	encodes the 117 amino acid sequence given in ABR06244; and (5) use of
CC	the polynucleotide for diagnosis and/or treatment of diseases caused by
CC	abnormal positioning of neural cells, and stimulation of association of
CC	reelin protein. The mouse reelin protein CR-50 epitope region has
CC	neuroprotective activity, and can be used in the diagnosis and treatment
CC	of cerebral disturbance due to an abnormal reelin gene and positioning
CC	of neurons.
XX	
SQ	Sequence 351 BP; 86 A; 98 C; 92 G; 75 T; 0 other;
Alignment Scores:	
Pred. No.:	9.49e-72 Length: 351
Score:	625.00 Matches: 117
Percent Similarity:	100.00% Conservative: 0
Best Local Similarity:	100.00% Mismatches: 0
Query Match:	100.00% Indels: 0
DB:	24 Gaps: 0
US-09-897-438B-2 (1-117) x ABL40165 (1-351)	
QY	1 GluGlnCysGlyThrIleMethHisGlyAsnAlaValThrPheCysGluProTyrGlyPro 20
Dd	
Dd	1 GACAGTGTTGGCACCATCATGCATGCAATGTCTACCTTCGTGAGCGGTACGCCCT 60
QY	21 ArgGluLeuThrThrThrCysLeuAsnThrThrThrAlaserValLeuGlnPheSerIle 40
Dd	
Dd	61 CGAGAGCTGACCACCCACATGCCTGAACAACAACAGCATCTGTCTCCAGTTTTCCATT 120
QY	41 GlySerGlySercysArgPheSerTyrSerAspProSerIleThrValserTyrAlalys 60
Dd	
Dd	121 GGCTCAGGATCATGTCGATTTAGTTACTCTGACCCAGCATCACGTGTCTATACGCCAAG 180
QY	61 AsnAsnThrAlaAspTrpIleGlnLeuGluLysIleArgAlaProSerAsnValserThr 80
Dd	
Dd	181 AACAAATACCGTGATTTGGATTTCAGCTGGAGAATAATAGAGCCCCCTTCCAATGTGAGCACA 240
QY	81 ValIleHisIleLeuTyrLeuProGluGluAlaLysGlyGluSerValGlnPheGlnTrp 100
Dd	
Dd	241 GTCATCCACATCTGTGTACCTCCCAGGAAGCAAAGGGGAGAGGCGTGCAGTTCAGTGG 300
QY	101 LysGlnAspSerLeuArgvalGlyGluValTyrGluAlaCysTrpAlaLeu 117
Dd	
Dd	301 AAACAGCAGCTCGAGTGGGTGAGTGTATGAGGCTGTGTGGGCCCTG 351

CC acids at the C terminus, and a series of eight internal repeats of  
 CC 350-390 amino acids, each repeat containing two related sub-domains  
 CC that flank a pattern of conserved cysteine residues known as an  
 CC epidermal growth factor (EGF)-like motif. (I) has neuroprotective,  
 CC neurotropic and antilipemic activities, and can be used as a modulator  
 CC of reelin-LDLR interaction. (I) is useful in screen for compounds that  
 CC modulate reelin binding to an LDLR, in an assay system, where the assay  
 CC system comprises a microplate array and an automated robotic  
 CC microprocessor controlled system for adding and removing reagents to  
 CC the microplate array. The compounds identified by the above screening  
 CC method are useful as therapeutic agents to provide or alleviate a  
 CC diverse spectrum of diseases including neurodegenerative disorders such  
 CC as Alzheimer's disease, to facilitate neuronal regeneration after  
 CC injury, to prevent or alleviate lipid metabolism diseases, to enhance  
 CC cognitive functions and memory or to ameliorate other developmental  
 CC disorders. The present sequence encodes mouse (Mus musculus) reelin,  
 CC which is used in the exemplification of the present invention.  
 XX  
 SQ Sequence 11673 BP; 2831 A; 2985 C; 2985 G; 2872 T; 0 other;

Alignment Scores:  
 Pred. No.: 1.37e-69 Length: 11673  
 Score: 625.00 Matches: 117  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 24 Gaps: 0

US-09-897-438b-2 (1-117) x ABA92603 (1-11673)

QY 1 GluGlnCysGlyThrIleMetHisGlyAsnAlaValThrPheCysGluProTyrGlyPro 20  
 DB 970 GAGCAGTGTGGACCATCATGCGAATGCTGTACCTTCTGTGAGCCGTACGGCCCT 1029

QY 21 ArgGluLeuThrThrCysLeuAsnThrThrAlaSerValLeuGlnPheSerIle 40  
 DB 1030 CGAGAGCTGACCAACACATGCTGACACAAACAGACATCTGTCTCCAGTTTCCATT 1089

QY 41 GlySerGlySerCysArgPheSerTyrSerAspProSerIleThrValSerTyrAlaLys 60  
 DB 1090 GGGTCAGGATCATGTCGATTAGTTACTCTGACCCAGCATCTGTGTATACGCCAAG 1149

QY 61 AsnAsnThrAlaAspTrpIleGlnLeuLysIleArgAlaProSerAsnValSerThr 80  
 DB 1150 AACAAATACCGCTGATTCGATTCAGCTGAGAAAATAGACCCCTTCCATGTGAGCACA 1209

QY 81 ValIleHisIleLeuTyrLeuProGluGluAlaLysGlyGluSerValGlnPheGlnTrp 100  
 DB 1210 GTCATCCACATCTCTGTACCTCCCGAGGAAAGCCAAAGGGAGAGCGTGCAGTTCAGTGG 1269

QY 101 LysGlnAspSerLeuArgValGlyGluValTyrGluAlaCysTrpAlaLeu 117  
 DB 1270 AACAGGACAGCCCTGCGAGTGGTGTATGAGGCTGTCTGGGCCCTG 1320

RESULT 4  
 ABI99284  
 ID ABI99284 standard; cDNA; 11673 BP.  
 AC ABI99284;  
 XX  
 XX 07-MAR-2002 (first entry)  
 XX  
 DE Mouse ischaemic condition related cDNA sequence SEQ ID NO:128.  
 XX  
 KW Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;  
 KW vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.  
 XX  
 OS Mus musculus.  
 XX  
 PN W0200188188-A2.  
 XX  
 PD 22-NOV-2001.  
 XX

QY 1 GluGlnCysGlyThrIleMetHisGlyAsnAlaValThrPheCysGluProTyrGlyPro 20  
 DB 970 GAGCAGTGTGGACCATCATGCGAATGCTGTACCTTCTGTGAGCCGTACGGCCCT 1029

QY 21 ArgGluLeuThrThrCysLeuAsnThrThrAlaSerValLeuGlnPheSerIle 40  
 DB 1030 CGAGAGCTGACCAACACATGCTGACACAAACAGACATCTGTCTCCAGTTTCCATT 1089

QY 41 GlySerGlySerCysArgPheSerTyrSerAspProSerIleThrValSerTyrAlaLys 60  
 DB 1090 GGGTCAGGATCATGTCGATTAGTTACTCTGACCCAGCATCTGTGTATACGCCAAG 1149

QY 61 AsnAsnThrAlaAspTrpIleGlnLeuLysIleArgAlaProSerAsnValSerThr 80  
 DB 1150 AACAAATACCGCTGATTCGATTCAGCTGAGAAAATAGACCCCTTCCATGTGAGCACA 1209

QY 81 ValIleHisIleLeuTyrLeuProGluGluAlaLysGlyGluSerValGlnPheGlnTrp 100  
 DB 1210 GTCATCCACATCTCTGTACCTCCCGAGGAAAGCCAAAGGGAGAGCGTGCAGTTCAGTGG 1269

QY 101 LysGlnAspSerLeuArgValGlyGluValTyrGluAlaCysTrpAlaLeu 117  
 DB 1270 AACAGGACAGCCCTGCGAGTGGTGTATGAGGCTGTCTGGGCCCTG 1320

RESULT 3  
 ABA92603  
 ID ABA92603 standard; cDNA; 11673 BP.  
 AC ABA92603;  
 XX  
 XX 21-MAR-2002 (first entry)  
 XX  
 DE Mouse reelin encoding cDNA SEQ ID NO:3.  
 XX  
 KW Mouse; reelin; low density lipoprotein receptor; LDLR; neuroprotective;  
 KW extracellular glycoprotein; neurotropic; antilipemic; Alzheimer's disease;  
 KW neurodegenerative disorder; neuronal regeneration; cognitive function;  
 KW lipid metabolism disease; memory; developmental disorder; gene; ss.  
 XX  
 OS Mus musculus.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 283..10668  
 FT /\*tag= a  
 FT /product= "mouse reelin"  
 XX  
 PN US6323177-B1.  
 XX  
 PD 27-NOV-2001.  
 XX  
 PF 16-JUN-1999; 99US-0334220.  
 XX  
 PR 16-JUN-1999; 99US-0334220.  
 XX  
 XX (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.  
 XX  
 XX Curran T, D'Arcangelo G;  
 XX  
 DR WPI; 2002-096596/13.  
 DR P-PSDB; ABB05008.  
 XX  
 XX Novel composition useful for screening compounds that modulate Reelin  
 PT binding to low density lipoprotein receptor, comprising an isolated  
 PT Reelin polypeptide and low density lipoprotein receptor -  
 XX  
 PS Example 1; Column 65-74; 45pp; English.  
 XX  
 CC The present invention describes a composition (I) comprising an  
 CC isolated reelin protein (ii) bound to an isolated low density lipoprotein  
 CC receptor (LDLR) (iii). (ii) is an extracellular glycoprotein of  
 CC approximately 385 kDa containing a small region of similarity with  
 CC F-spondin at the N terminus, a stretch of positively charged amino

PF 18-MAY-2001; 2001WO-JP04192.  
XX  
PR 18-MAY-2000; 2000JP-0145977.  
XX  
XX (UYN1-) UNIV NIHON SCHOOL JURIDICAL PERSON.  
XX  
PI Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;  
XX  
XX WPI; 2002-034733/04.  
DR - P-PSDB; ABB57065.  
XX  
XX Examining the ischemic condition (e.g. occlusive ischemia) by measuring  
PT expression levels of particular genes defined in the specification or  
PT by determining the expression profile of a gene group comprising these  
PT genes -  
XX  
PS Claim 2; Page 362-385; 2690pp; English.  
XX  
XX The present invention describes a method for examining ischaemic  
CC conditions, comprising measuring the expression levels of particular  
CC genes (I) in a test sample or determining the expression profile of a  
CC gene group in the sample comprising genes selected from (I). The method  
CC is useful for examining the ischaemic condition (e.g. compressive  
CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring  
CC expression levels of particular genes (ABI99202 to ABI9912, encoding  
CC the protein sequences in ABB57020 to ABB57374) or by determining the  
CC expression profile of a gene group comprising these genes. The  
CC expression levels or expression profiles produced by these genes are  
CC used as an indicator when screening for ischaemic condition-improving  
CC drugs or therapeutics for ischaemic diseases. ABI9913 and ABI9914  
CC represent PCR primers for a mouse ischaemic condition related sequence,  
CC which are used in the exemplification of the present invention.  
XX  
SQ Sequence 11673 BP; 2831 A; 2985 C; 2985 G; 2872 T; 0 other;

Alignment Scores:  
Pred. No.: 1.37e-69 Length: 11673  
Score: 625.00 Matches: 117  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 24 Gaps: 0  
US-09-897-438B-2 (1-117) x ABI99284 (1-11673)  
Qy 1 GluGlnCysGlyThrIleMetHisGlyAsnAlaValThrPheCysGluProTyrGlyPro 20  
Db 970 GAGCAGTGTGCACCATCATGCGAATGCTGTACCTTCTGTGAGCGGTACGGCCCT 1029  
Qy 21 ArgGluLeuThrThrCysLeuAsnThrThrThrAlaSerValLeuGlnPheSerile 40  
Db 1030 CGAGAGCTGACCACCATGCTGAACACAAACAGCATGTGCTCCAGTTTTCATT 1089  
Qy 41 GlySerGlySerCysArgPheSerTyrSerAspProSerIleThrValSerTyrAlaLys 60  
Db 1090 GGGTCAGATCATTCGATTTAGTTACTGTGACCCAGCATCATCTGTGTATAGCCCAAG 1149  
Qy 61 AsnAsnThrAlaAspTptIleGlnLeuGluLysIleArgAlaProSerAsnValSerThr 80  
Db 1150 AACAAATACCGCTGATTTGGATTACCTCGAGAAATTAAGACCCCTTCCAATGTGAGCACA 1209  
Qy 81 ValIleHisIleLeuTyrLeuProGluGluAlaLysGlyGluSerValGlnPheGlnTrp 100  
Db 1210 GTCATCCACATCCCTGTACCTCCCGAGGAGGCCAAAGCGGAGCGGTCCAGTGG 1269  
Qy 101 LysGlnAspSerLeuArgValGlyGluValTyrGluAlaCysTrpAlaLeu 117  
Db 1270 AACAGGACAGCCTGCGAGTGGGTGAGGTGTATGAGGCCCTGCTGGGCCCTG 1320  
RESULT 5  
ABN96939  
ID ABN96939 standard; DNA; 11580 BP.  
XX

AC ABN96939;  
XX  
DT 13-AUG-2002 (first entry)  
XX  
DE Gene #3437 used to diagnose liver cancer.  
XX  
KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;  
KW metastatic liver tumour; cytostatic; expression profile; disease state;  
KW disease progression; drug toxicity; drug efficacy; drug metabolism.  
XX  
OS Homo sapiens.  
XX  
PN WO200229103-A2.  
XX  
PD 11-APR-2002.  
XX  
XX 02-OCT-2001; 2001WO-US30589.  
XX  
XX 02-OCT-2000; 2000US-237054P.  
XX  
XX (GENE-) GENE LOGIC INC.  
XX  
PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;  
XX  
DR WPI; 2002-426119/45.  
XX  
XX Diagnosing and detecting the progression of liver cancer,  
XX hepatocellular carcinoma or metastatic liver tumor in a patient,  
XX involves detecting the level of expression of two or more genes in a  
XX liver tissue sample -  
XX  
XX Claim 1; SEQ ID NO 3437; 298pp; English.

XX The invention relates to a novel method for diagnosing and detecting the  
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver  
CC tumour in a patient, and differentiating metastatic liver cancer from  
CC hepatocellular carcinoma in a patient, involving detecting the level of  
CC expression of two or more genes represented in ABN97455 in a  
CC tissue sample. The method of the invention has hepatotropic, and  
CC cytostatic activity. The method is useful for diagnosing and detecting  
CC the progression of liver cancer, hepatocellular carcinoma and metastatic  
CC liver carcinoma in a patient. The method is useful for identifying  
CC expression profiles which serve as useful diagnostic markers as well as  
CC markers that can be used to monitor disease states, disease progression,  
CC drug toxicity, drug efficacy and drug metabolism  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 11580 BP; 3014 A; 2696 C; 2753 G; 3116 T; 1 other;

Alignment Scores:  
Pred. No.: 1e-62 Length: 11580  
Score: 572.00 Matches: 106  
Percent Similarity: 95.73% Conservative: 6  
Best Local Similarity: 90.60% Mismatches: 5  
Query Match: 91.52% Indels: 0  
DB: 24 Gaps: 0  
US-09-897-438B-2 (1-117) x ABN96939 (1-11580)  
Qy 1 GluGlnCysGlyThrIleMetHisGlyAsnAlaValThrPheCysGluProTyrGlyPro 20  
Db 860 GAACAGTGTGGCGGATATGATGCGAATGCGGTACCTTCTGTGAAACATATGGCCCA 919  
Qy 21 ArgGluLeuThrThrCysLeuAsnThrThrThrAlaSerValLeuGlnPheSerile 40  
Db 920 CGAGAACTGATTACACAGCGCTTAATACAAACAGCTTCTGTCTCCCAATTTTCATT 979  
Qy 41 GlySerClySerCysArgPheSerTyrSerAspProSerIleThrValSerTyrAlaLys 60  
Db 980 GGGTCAGGTTTCATGTCGCTTAGTTATTACAGACCCAGCATCATCGTGTATATGCAAG 1039

QY 61 AsnAsnThrAlaAspTrpIleGlnLeuGluLysIleArgAlaProSerAsnValSerThr 80  
 Db 1040 AATAACTCTCGGACTGGATTACGTAGAGAAAATAGAGCCCTTCCAAATGTCACACA 1099  
 QY 81 ValIleHisIleLeuTyrLeuProGluGluAlaLysGlyGluSerValGlnPheGlnTrp 100  
 Db 1100 ATCATCCATATCTCTACCTTCTGAGGACGCCAAAGGGGAGAAATGCCAATTTCAAGTG 1159  
 QY 101 LysGlnAspSerLeuArgValGlyGluValTyrGluAlaCysTrpAlaLeu 117  
 Db 1160 AAGCAGGAAAATCTCTGTAGGTGAAGTGTATGAAGCCCTGTGGGCCCTTA 1210  
 RESULT 6  
 ID ABA92604  
 AC ABA92604  
 XX ABA92604 standard; cdna; 11580 BP.  
 XX  
 XX 21-MAR-2002 (first entry)  
 XX  
 DE Human reelin encoding cdna SEQ ID NO:4.  
 DE  
 KW Human; reelin; low density lipoprotein receptor; LDLR; neuroprotective;  
 KW extracellular glycoprotein; neotropic; antilipaeamic; Alzheimer's disease;  
 KW neurodegenerative disorder; neuronal regeneration; cognitive function;  
 KW lipid metabolism disease; memory; developmental disorder; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 176..10558  
 FT /\*tag= a  
 FT /\*product= "human reelin"  
 XX  
 PN US62323177-B1.  
 PD 27-NOV-2001.  
 PF 16-JUN-1999; 99US-0334220.  
 XX 16-JUN-1999; 99US-0334220.  
 PR  
 XX (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.  
 XX  
 PI Curran T, D'Arcangelo G;  
 XX  
 DR WPI; 2002-096596/13.  
 DR P-PSDB; ABB05007.  
 XX  
 PT Novel composition useful for screening compounds that modulate Reelin  
 PT binding to low density lipoprotein receptor, comprising an isolated  
 PT Reelin polypeptide and low density lipoprotein receptor -  
 XX  
 PS Example 1; Column 75-84; 45pp; English.  
 XX  
 CC The present invention describes a composition (I) comprising an  
 CC isolated reelin protein (II) bound to an isolated low density lipoprotein  
 CC receptor (LDLR) (III). (II) is an extracellular glycoprotein of  
 CC approximately 385 kda containing a small region of similarity with  
 CC F-spondin at the N terminus, a stretch of positively charged amino  
 CC acids at the C terminus, and a series of eight internal repeats of  
 CC 350-390 amino acids, each repeat containing two related sub-domains  
 CC that flank a pattern of conserved cysteine residues known as an  
 CC epidermal growth factor (EGF)-like motif. (I) has neuroprotective,  
 CC neotropic and antilipaeamic activities, and can be used as a modulator  
 CC of reelin-LDLR interaction. (I) is useful in screen for compounds that  
 CC modulate reelin binding to an LDLR, in an assay system, where the assay  
 CC system comprises a microplate array and an automated robotic  
 CC microprocessor controlled system for adding and removing reagents to  
 CC the microplate array. The compounds identified by the above screening  
 CC method are useful as therapeutic agents to provide or alleviate a  
 CC diverse spectrum of diseases including neurodegenerative disorders such  
 CC as Alzheimer's disease, to facilitate neuronal regeneration after

CC injury, to prevent or alleviate lipid metabolism diseases, to enhance  
 CC cognitive functions and memory or to ameliorate other developmental  
 CC disorders. The present sequence encodes human reelin, which is used in  
 CC the exemplification of the present invention.  
 SQ Sequence 11580 BP; 3014 A; 2696 C; 2753 G; 3116 T; 1 other;  
 XX  
 Alignment Scores:  
 Pred. No.: 1e-62 Length: 11580  
 Score: 572.00 Matches: 106  
 Percent Similarity: 95.73% Conservative: 6  
 Best Local Similarity: 90.60% Mismatches: 5  
 Query Match: 91.52% Indels: 0  
 DB: 24 Gaps: 0  
 US-09-897-438B-2 (1-117) x ABA92604 (1-11580)  
 QY 1 GluGlnCysGlyThrIleMetHisGlyAsnAlaValThrPheCysGluProTyrGlyPro 20  
 Db 860 GAACAGTGTGGCGGATTTATGCGAATGGCAATGCCCTTCTGTGAACCATATGCCCA 919  
 QY 21 ArgGluLeuThrThrCysLeuAsnThrThrThrAlaSerValLeuGlnPheSerIle 40  
 Db 920 CGAGAACTCATTAACACAGGCCCTTAATACAACACAGCTTCTGCTCCCAATTTCCATT 979  
 QY 41 GlySerGlySerCysArgPheSerTyrSerAspProSerIleThrValSerTyrAlaLys 60  
 Db 980 GGGTCAGGTTCATGTCGCTTAGTTATTAGACCCAGCATCATCGTTATATATGCCAAG 1039  
 QY 61 AsnAsnThrAlaAspTrpIleGlnLeuGluLysIleArgAlaProSerAsnValSerThr 80  
 Db 1040 AATAACTCTCGGACTGGATTACGTAGAGAAAATAGAGCCCTTCCAAATGTCACACA 1099  
 QY 81 ValIleHisIleLeuTyrLeuProGluGluAlaLysGlyGluSerValGlnPheGlnTrp 100  
 Db 1100 ATCATCCATATCTCTACCTTCTGAGGACGCCAAAGGGGAGAAATGCCAATTTCAAGTG 1159  
 QY 101 LysGlnAspSerLeuArgValGlyGluValTyrGluAlaCysTrpAlaLeu 117  
 Db 1160 AAGCAGGAAAATCTCTGTAGGTGAAGTGTATGAAGCCCTGTGGGCCCTTA 1210  
 RESULT 7  
 ID AAS89484  
 AC AAS89484 standard; cdna; 11632 BP.  
 XX  
 AC AAS89484;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE DNA encoding novel human diagnostic protein #25288.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 XX 23-AUG-2000; 2000US-0649167.  
 XX (HYSE-) HYSEQ INC.  
 XX Drmanac RT, Liu C, Tang YT;  
 XX WPI; 2001-639362/73.  
 DR P-PSDB; ABB25297.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
PS Claim 1; SEQ ID No 25288; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 11632 BP; 3028 A; 2716 C; 2770 G; 3117 T; 1 other;

## Alignment Scores:

Pred. No.: 1.01e-62 Length: 11632  
Score: 572.00 Matches: 106  
Percent Similarity: 95.73% Conservative: 6  
Best Local Similarity: 90.60% Mismatches: 5  
Query Match: 91.52% Indels: 0  
DB: 23 Gaps: 0

US-09-897-438B-2 (1-117) x AAS99484 (1-11632)

Qy 1 GluInCysGlyThrIleMetHisGlyAsnAlaValThrPheCysGluProTyrGlyPro 20  
Db 860 GAACAGTGTGGCGGATATGCAATGGCAATGCCGTACCTCTCTGGAACCATATGGCCCA 919  
Qy 21 ArgGluLeuThrThrCysLeuAsnThrThrThrAlaSerValLeuGlnPheSerIle 40  
Db 920 CGAGAACTGATTACACAGGCGCTTAATACACACAGCTTCTGCTCCCAATTTCCATT 979  
Qy 41 GlySerGlySerCysArgPheSerTyrSerAspProSerIleThrValSerTyrAlaLys 60  
Db 980 GGGTCAGGTTTCATCGCTTAGTATTACAGACCCCGCATCATCGTGTATATGCCAAG 1039  
Qy 61 AsnAsnThrAlaAspTrpIleGlnLeuGluLysIleArgAlaProSerAsnValSerThr 80  
Db 1040 AATAACTCTCGGACTGGATTCAGCTAGAGAAATTAAGACCCCTTCCATGTCAGCACA 1099  
Qy 81 ValIleHisIleLeuTyrLeuProGluAlaLysGlyGluSerValGlnPheGlnTrp 100  
Db 1100 ATCATCCATATCTCTCTACCTCTCTGAGGACGCCAAGGGAGAATGTCGAATTTCACTGG 1159  
Qy 101 LysGlnAspSerLeuArgValGlyGluValTyrGluAlaCysTrpAlaLeu 117  
Db 1160 AAGCAGAAATCTTCGTGAGGTGAAGTGTATGAAGCCCTGCTGGGCCCTTA 1210  
RESULT 8  
AAD22779  
ID AAD22779 standard; cDNA; 2025 BP.  
XX  
AC AAD22779;  
XX  
DT 26-FEB-2002 (first entry)  
XX  
DE Xenopus laevis intact reelin cDNA.

XX Frog; reelin; F-spondin domain; CR-50 epitope; gene therapy; agyria;  
KW polymicrogyria; ectopic gray matter; ss.  
OS  
XX Xenopus laevis.

Key Location/Qualifiers  
CDS 157..2025  
FT /\*tag= a  
FT /product= "Xenopus intact reelin protein"  
FT /transl\_except= (pos:1606..1608, aa:Cys)  
FT /transl\_except= (pos:1639..1641, aa:Cys)  
FT /note= "CDS does not include stop codon"  
FT /partial

EPI149844-A2.

31-OCT-2001.

11-APR-2001; 2001EP-0303411.

11-APR-2000; 2000JP-0109954.

(RIKE ) RIKEN KK.

Mikoshiha K, Tabata H, Nakajima K;

WPI: 2002-019320/03.

P-PSDB; AAE13607.

XX Novel truncated Reelin protein containing F-spondin domain and CR-50  
PT recognition site of Reelin protein, but not having Reelin repeat site,  
PT useful to treat diseases including agyria due to abnormal neuron  
PT alignment

Example 1; Fig 1A; 47pp; English.

XX The invention relates to a truncated Reelin protein comprising a  
CC F-spondin domain and a CR-50 recognition site but no reelin protein  
CC repeat site. Reelin is an essential molecule in developing a normal  
CC laminated structure of cerebrum. The truncated reelin protein and its  
CC DNA are useful for treating diseases including agyria, polymicrogyria,  
CC and ectopic gray matter due to abnormal neuronal alignment. Truncated  
CC reelin protein DNA is useful in gene therapy. The present sequence is  
CC a cDNA encoding Xenopus laevis intact reelin protein.

SQ Sequence 2025 BP; 541 A; 419 C; 460 G; 604 T; 1 other;

## Alignment Scores:

Pred. No.: 2.47e-54 Length: 2025  
Score: 499.00 Matches: 87  
Percent Similarity: 88.89% Conservative: 17  
Best Local Similarity: 74.36% Mismatches: 13  
Query Match: 79.84% Indels: 0  
DB: 24 Gaps: 0

US-09-897-438B-2 (1-117) x AAD22779 (1-2025)

Qy 1 GluInCysGlyThrIleMetHisGlyAsnAlaValThrPheCysGluProTyrGlyPro 20  
Db 847 GAGCAGTGTGTCGAATTTATGCAATGGTGGGCGAGTCATCTTTTGTGATCGGTATGGACCA 906  
Qy 21 ArgGluLeuThrThrCysLeuAsnThrThrThrAlaSerValLeuGlnPheSerIle 40  
Db 907 AGAGAATTGATAACTGTTCAATTCACACACAACTACGGCATCTGTTTGCAGTTCATT 966  
Qy 41 GlySerGlySerCysArgPheSerTyrSerAspProSerIleThrValSerTyrAlaLys 60  
Db 967 GGTCAGGATCGTCAGGTTTCAGTATTACAGACCCCTGGAATTTGTTGTGTCATACACAAG 1026  
Qy 61 AsnAsnThrAlaAspTrpIleGlnLeuGluLysIleArgAlaProSerAsnValSerThr 80  
Db 1027 AATAATTCATCACTAGTGGATGCCATTGGAGAGAATTAGTCTCTCTCCAAATGTAGCACC 1086

QY 81 ValIleHisIleLeuTyrLeuProGluGluAlaLysGlyGluSerValGlnPheGlnTrp 100  
 ::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
 Db 1087 ATCATTACATATTATACCTACCTCTGAAGCTTAAGGAGAAATGTGAATTCGGTTGG 1146

QY 101 LysGlnAspSerLeuArgValGlyGluValTyrGluAlaCysTrpAlaLeu 117  
 ::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
 Db 1147 AGGCAGGAGAACATGCAGGCAGGTGATGTGTATGAAGCCTGCTGGGCACCTG 1197

RESULT 9  
 AAD22753  
 ID AAD22753 standard; cDNA; 2274 BP.  
 XX  
 AC AAD22753;  
 XX  
 DT 26-FEB-2002 (first entry)  
 XX  
 DE Xenopus laevis truncated reelin cDNA.  
 XX  
 KW Frog; reelin; F-spondin domain; CR-50 epitope; gene therapy; agyria;  
 KW polymicrogyria; ectopic gray matter; ss.  
 XX  
 OS Xenopus laevis.  
 XX

Key Location/Qualifiers  
 FH 157..1455  
 FT CDS  
 FT /tag= a  
 FT /product= "X. laevis truncated reelin protein"  
 FT 157..234  
 FT sig\_peptide  
 FT /tag= b  
 FT 235..1452  
 FT mat\_peptide  
 FT /tag= c  
 FT /product= "Mature truncated reelin protein"  
 FT 241..726  
 FT misc\_feature  
 FT /tag= d  
 FT /note= "Encodes F-spondin domain"  
 FT 847..1197  
 FT misc\_feature  
 FT /tag= e  
 FT /note= "Encodes CR-50 epitope region"  
 FT 2254..2259  
 FT polyA\_signal  
 FT /tag= f

XX EP1149844-A2.  
 XX  
 XX 31-OCT-2001.  
 XX  
 PF 11-APR-2001; 2001EP-0303411.  
 XX  
 PR 11-APR-2000; 2000JP-0109954.  
 XX  
 PA (RIKE ) RIKEN KK.  
 XX  
 PI Mikoshiba K, Tabata H, Nakajima K;  
 XX  
 DR WPI; 2002-019320/03.  
 DR P-PSDB; AAEL3605.  
 XX  
 PT Novel truncated Reelin protein containing F-spondin domain and CR-50  
 PT recognition site of Reelin protein, but not having Reelin repeat site,  
 PT useful to treat diseases including agyria due to abnormal neuron  
 PT alignment -  
 XX  
 PS Claim 8; Page 11-16; 47pp; English.  
 XX  
 CC The invention relates to a truncated Reelin protein comprising a  
 CC F-spondin domain and a CR-50 recognition site but no reelin protein  
 CC repeat site. Reelin is an essential molecule in developing a normal  
 CC laminated structure of cerebrum. The truncated reelin protein and its  
 CC DNA are useful for treating diseases including agyria, polymicrogyria,  
 CC and ectopic gray matter due to abnormal neuronal alignment. Truncated  
 CC reelin protein DNA is useful in gene therapy. The present sequence is  
 CC a cDNA encoding Xenopus laevis truncated reelin protein.  
 XX

SQ Sequence 2274 BP; 660 A; 450 C; 469 G; 694 T; 1 other;  
 Alignment Scores:  
 Pred. No.: 2,91e-54 Length: 2274  
 Score: 499.00 Matches: 87  
 Percent Similarity: 88.89% Conservative: 17  
 Best Local Similarity: 74.36% Mismatches: 13  
 Query Match: 79.84% Indels: 0  
 DB: 24 Gaps: 0

US-09-897-438B-2 (1-117) x AAD22753 (1-2274)

QY 1 GluGlnCysGlyThrIleMetHisGlyAsnAlaValThrPheCysGluProTyrGlyPro 20  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 Db 847 GAGCAGTGTGTCATATTATGTCATGTTGGGCGAGTCACCTTTTGTGATCCGTATGGACCA 906

QY 21 ArgGluLeuThrThrThrCysLeuAsnThrThrThrAlaSerValLeuGlnPheSerIle 40  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 Db 907 AGAGAATTGATAACTGTTCAAATGAACACAACTACGGCATCTGTTTGCAGTTTCTATT 966

QY 41 GlySerGlySerCysArgPheSerTyrSerAspProSerIleThrValSerTyrAlaLys 60  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 Db 967 GGCTCAGGATCGTGCGAGGTTTCAGCTATTCAGACCCCTGGAATTTGTGTGCATACACAAAG 1026

QY 61 AsnAsnThrAlaAspTrpIleGlnLeuGluLysIleArgAlaProSerAsnValSerThr 80  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 Db 1027 AATAATTCATCTAGTTGGATGCCATTTGGAGAGAATTAGTCTCTCTCCATGTTAGCACC 1086

QY 81 ValIleHisIleLeuTyrLeuProGluGluAlaLysGlyGluSerValGlnPheGlnTrp 100  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 Db 1087 ATCATTACATATTATTTACCTACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1146

QY 101 LysGlnAspSerLeuArgValGlyGluValTyrGluAlaCysTrpAlaLeu 117  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 Db 1147 AGCAGGAGAACATGCAGGCAGGTGATGTGTATGAAGCCTGCTGGGCACCTG 1197

RESULT 10  
 AAK07720/C  
 ID AAK07720 standard; DNA; 413 BP.  
 XX  
 AC AAK07720;  
 XX  
 DT 05-NOV-2001 (first entry)  
 XX  
 DE Human brain expressed single exon probe SEQ ID NO: 7711.  
 XX  
 KW Human; brain expressed exon; gene expression analysis; probe;  
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
 KW epilepsy; cancer; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157275-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00667.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-483446/52.  
 XX  
 PT Single exon nucleic acid probes for analyzing gene expression in human

```

PT brains -
XX Example 4; SEQ ID NO: 7711; 650pp + Sequence Listing; English.
PS
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX
SQ Sequence 413 BP; 134 A; 86 C; 67 G; 126 T; 0 other;

Alignment Scores:
Pred. No.: 2.74e-09 Length: 413
Score: 144.00 Matches: 36
Percent Similarity: 47.17% Conservative: 14
Best Local Similarity: 33.96% Mismatches: 24
Query Match: 23.04% Indels: 32
DB: 22 Gaps: 4

US-09-897-438B-2 (1-117) x AAK07720 (1-413)
QY 1 GluGlnCysGlyThrIleMethHisGlyAsnAlaValThrPheCysGluProTyrGlyPro 20
DB 323 GAACAGTGTGGCGGCGATTATGCATGGCAATGCCGTTCACCTTCTGTGAACCATATGGCCCA 264
QY 21 ArgGluLeuThrThrCysLeuAsnThrThrAlaSerValLeuGlnPheSerile 40
DB 263 CGAGAACTGGTAAGTATGCTATTCATGTA-----TTATTCATTATAGTATT 213
QY 41 GlySerGlySerCysArgPheSerTyr-----SerAspProSerile 54
DB 212 TACTGAGCTTTTGGCAACAAAAGTTTCTTATGCTCTTAAAGTTCAGATGAAAT--- 156
QY 55 ThrValSerTyrAlaLysAsnThrAlaAspTrpIleGlnLeuGluLysIleArgAla 74
DB 156 ----- 156
QY 75 ProSerAsnValSerThrValIleHisIleLeuTyrLeuProGluGluAlaLysGlyGlu 94
DB 155 ---TCAATATGATTATGAGTCTCTATTACAAATATCTT---AAATTTGGGAGGGGTGAT 102
QY 95 SerValGlnPheGlnTrp 100
DB 101 TCTATGGCTATTTTGG 84

RESULT 11
AAK33553/c
ID AAK33553 standard; DNA; 413 BP.
XX
AC AAK33553;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 8110.
XX
DE Human; bone marrow expressed exon; gene expression analysis; probe;
XX
DE Microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.

```

```

PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -
XX
XX Example 4; SEQ ID NO: 8110; 658pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is one of
XX the probes of the invention.
XX
SQ Sequence 413 BP; 134 A; 86 C; 67 G; 126 T; 0 other;

Alignment Scores:
Pred. No.: 2.74e-09 Length: 413
Score: 144.00 Matches: 36
Percent Similarity: 47.17% Conservative: 14
Best Local Similarity: 33.96% Mismatches: 24
Query Match: 23.04% Indels: 32
DB: 22 Gaps: 4

US-09-897-438B-2 (1-117) x AAK33553 (1-413)
QY 1 GluGlnCysGlyThrIleMethHisGlyAsnAlaValThrPheCysGluProTyrGlyPro 20
DB 323 GAACAGTGTGGCGGCGATTATGCATGGCAATGCCGTTCACCTTCTGTGAACCATATGGCCCA 264
QY 21 ArgGluLeuThrThrCysLeuAsnThrThrAlaSerValLeuGlnPheSerile 40
DB 263 CGAGAACTGGTAAGTATGCTATTCATGTA-----TTATTCATTATAGTATT 213
QY 41 GlySerGlySerCysArgPheSerTyr-----SerAspProSerile 54
DB 212 TACTGAGCTTTTGGCAACAAAAGTTTCTTATGCTCTTAAAGTTCAGATGAAAT--- 156
QY 55 ThrValSerTyrAlaLysAsnThrAlaAspTrpIleGlnLeuGluLysIleArgAla 74
DB 156 ----- 156
QY 75 ProSerAsnValSerThrValIleHisIleLeuTyrLeuProGluGluAlaLysGlyGlu 94
DB 155 ---TCAATATGATTATGAGTCTCTATTACAAATATCTT---AAATTTGGGAGGGGTGAT 102
QY 95 SerValGlnPheGlnTrp 100
DB 101 TCTATGGCTATTTTGG 84

RESULT 12
ABS08422/c
ID ABS08422 standard; DNA; 413 BP.
XX
AC ABS08422;
XX
DT 19-AUG-2002 (first entry)
XX
DE Human genome-derived single exon probe from lung SEQ ID No 8413.
XX
DE Human; ds; single exon probe; asthma; lung cancer; COPD; IID;
XX
DE Chronic obstructive pulmonary disease; interstitial lung disease;
XX
DE familial idiopathic pulmonary fibrosis; neurofibromatosis;
XX
DE tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
XX
DE Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;

```





XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 4576 BP; 1222 A; 1023 C; 1109 G; 1222 T; 0 other;

Alignment Scores:  
Pred. No.: 0.00246 Length: 4576  
Score: 109.50 Matches: 37  
Percent Similarity: 48.36% Conservative: 22  
Best Local Similarity: 30.33% Mismatches: 52  
Query Match: 17.52% Indels: 11  
DB: 23 Gaps: 5

US-09-897-438B-2 (1-117) x AAS89475 (1-4576)

Qy 3 CysGlyThrIleMetHisGlyAsnAlaValThrPheCysGluProTyrGlyProArgGlu 22  
Db 2855 TGTGGAACACTTAACGATGGCAATCTCTACTTCAATGGCCCT---GGGAAAGGGAA 2911  
Qy 23 LeuThrThrCysLeuAsnThrThrAlaSerValLeuGlnPhe-----SerIleThr 40  
Db 2912 GCCGGACGGTCCCTCTCGACACCGAGGAATATCAGACTTGTCAATTTATATACAAAT 2971  
Qy 41 GlySerGlySerCysArgPheSerTyrSerAspPro-----SerIleThr 55  
Db 2972 GGAACAAACTTACAGGATACCTGCATCAACCAAGAACTAGAAATGAGGCTTATT 3031  
Qy 56 ValSerTyrAlaLysAsnThrAlaAspTrpIleGlnLeuGluLysIleArgAlaPro 75  
Db 3032 GTTCAGTATTCAATGACAATGGGATCTCTGGCATTTGCTTCGAGAGTTGGACTTCATG 3091  
Qy 76 SerAsnValSerThrValIleHisIleLeuTyrLeuProGluGluAlaLysGlyCysSer 95  
Db 3092 TCCCTTCGGAAACACACATCATTTCCATTGACCTGCCACAGGACGGGAACACCTGCA 3151  
Qy 96 ValGlnPheGlnTrpLysGlnAspSerLeuArgValGlyGluValTyrGluAlaCysTrp 115  
Db 3152 ACGGATTTTCGATGGTGCACCG-----CAACATGGGAG---CATTGACCCCATGG 3202  
Qy 116 AlaLeu 117  
Db 3203 GCCTTG 3208  
RESULT 14  
AAK20384/C  
ID AAK20384 standard; DNA; 87 BP.  
XX AC AAK20384;  
XX AC AAK20384;  
XX 05-NOV-2001 (first entry)  
DT Human brain expressed single exon probe SEQ ID NO: 20375.  
DE  
XX

KW Human; brain expressed exon; gene expression analysis; probe;  
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
KW epilepsy; cancer; ss.  
XX Homo sapiens.  
XX OS  
XX PN WO200157275-A2.  
XX PD 09-AUG-2001.  
XX PF 30-JAN-2001; 2001WO-US00667.  
XX PR 04-FEB-2000; 2000US-0180312.  
XX PR 26-MAY-2000; 2000US-0207456.  
XX PR 30-JUN-2000; 2000US-0608408.  
XX PR 03-AUG-2000; 2000US-0623366.  
XX PR 21-SEP-2000; 2000US-0234687.  
XX PR 27-SEP-2000; 2000US-0236359.  
XX PR 04-OCT-2000; 2000GB-0024263.  
XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-483446/52.  
XX Single exon nucleic acid probes for analyzing gene expression in human  
XX brains -  
XX Example 4; SEQ ID NO: 20375; 650pp + Sequence Listing; English.  
XX The present invention provides a number of single exon nucleic acid  
XX probes which are derived from genomic sequences expressed in the human  
XX brain. They can be used to measure gene expression in brain cell samples,  
XX which may enable the diagnosis and improved treatment of nervous system  
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
XX epilepsy and cancers. The present sequence is one of the probes of the  
XX invention.  
XX SQ Sequence 87 BP; 21 A; 25 C; 19 G; 22 T; 0 other;

Alignment Scores:  
Pred. No.: 1.88e-05 Length: 87  
Score: 107.00 Matches: 18  
Percent Similarity: 94.74% Conservative: 0  
Best Local Similarity: 94.74% Mismatches: 1  
Query Match: 17.12% Indels: 0  
DB: 22 Gaps: 0

US-09-897-438B-2 (1-117) x AAK20384 (1-87)

Qy 1 GluGlnCysGlyThrIleMetHisGlyAsnAlaValThrPheCysGluProTyrGly 19  
Db 59 GAACAGTGTGGCGGATATGTCATGGCAATGCCGTCACCTTCTGTGAACCATATGGC 3  
RESULT 15  
AAK46490/C  
ID AAK46490 standard; DNA; 87 BP.  
XX AC AAK46490;  
XX AC AAK46490;  
XX 06-NOV-2001 (first entry)  
DT Human bone marrow expressed single exon probe SEQ ID NO: 21047.  
DE  
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.  
XX OS Homo sapiens.  
XX PN WO200157276-A2.  
XX PD 09-AUG-2001.

```
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
XX Example 4; SEQ ID NO: 21047; 658pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
XX SQ Sequence 87 BP; 21 A; 25 C; 19 G; 22 T; 0 other;
XX
Alignment Scores:
Pred. No.: 1.88e-05 Length: 87
Score: 107.00 Matches: 18
Percent Similarity: 94.74% Conservative: 0
Best Local Similarity: 94.74% Mismatches: 1
Query Match: 17.12% Indels: 0
DB: 22 Gaps: 0
US-09-897-438B-2 (1-117) x AAK46490 (1-87)
QY 1 GluGlnCysGlyThrIleMetHisGlyAsnAlaValThrPheCysGluProTyrGly 19
Db 59 GAACAGTGTGGCGGATTATGTCATGGCAATGCCGTCACCTTCTGTGAACCATATGGC 3
Search completed: November 6, 2002, 19:30:02
Job time : 308 secs
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 6, 2002, 20:18:37 ; Search time 87 Seconds  
(without alignments)  
448.036 Million cell updates/sec

Title: US-09-897-438B-2

Perfect score: 625

Sequence: 1 ECGTTHGNAVTFCEPYGP.....FQWKQDSLVRGEVYEACWAL 117

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 310279 seqs, 166577418 residues

Total number of hits satisfying chosen parameters: 620558

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_plus\_p2n.model -DEV=xlip  
-Q=/cgn2\_1/US09897438/runat\_06112002\_101927\_11396/app\_query.fasta\_1.263  
-DB=Published\_Applications\_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bloms62  
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09897438 -CGN\_1\_1\_36 @runat\_06112002\_101927\_11396  
-NCPU=6 -ICPU=3 -NO\_XLPXY -NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=3 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published\_Applications\_NA.\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	625	100.0	351	10	US-09-897-438B-1 Sequence 1, Appli
2	572	91.5	11580	10	US-09-880-107-3436 Sequence 3436, Ap
3	224	35.8	499	10	US-09-864-761-14497 Sequence 14497, A
4	190	30.4	214	10	US-09-864-761-31037 Sequence 31037, A

C	5	144	23.0	413	10	US-09-864-761-11808	Sequence 11808, A
C	6	107	17.1	87	10	US-09-864-761-28379	Sequence 28379, A
C	7	72	11.5	795	10	US-09-974-300-2745	Sequence 2745, Ap
C	8	70	11.2	9274	10	US-09-885-535-3	Sequence 3, Appli
C	9	69	11.0	495	10	US-09-864-761-10582	Sequence 10582, A
C	10	68.5	11.0	468	10	US-09-917-800A-547	Sequence 547, App
C	11	68	10.9	978	10	US-09-886-055-2	Sequence 2, Appli
C	12	67	10.7	1017	10	US-09-802-371-3	Sequence 3, Appli
C	13	67	10.7	1585	10	US-09-802-371-1	Sequence 1, Appli
C	14	66.5	10.6	612	10	US-09-974-300-1207	Sequence 1207, Ap
C	15	66.5	10.6	2082	10	US-09-741-669-200	Sequence 200, App
C	16	66.5	10.6	143058	10	US-09-967-769A-316	Sequence 316, App
C	17	66	10.6	2001	10	US-09-841-683-3	Sequence 3, Appli
C	18	66	10.6	2466	10	US-09-922-138-4	Sequence 4, Appli
C	19	66	10.6	4237	10	US-09-764-877-3835	Sequence 3835, Ap
C	20	66	10.6	5504	8	US-08-913-322-1	Sequence 1, Appli
C	21	66	10.6	6124	8	US-08-913-322-21	Sequence 21, Appli
C	22	66	10.6	6124	10	US-09-967-769A-184	Sequence 184, App
C	23	66	10.6	6133	8	US-08-913-322-2	Sequence 2, Appli
C	24	66	10.6	6228	8	US-08-913-322-23	Sequence 23, Appli
C	25	65.5	10.5	2846	10	US-09-962-436-303	Sequence 303, App
C	26	65.5	10.5	2846	10	US-09-962-436-309	Sequence 309, App
C	27	65.5	10.5	5769	10	US-09-964-824A-284	Sequence 284, App
C	28	65	10.4	528	10	US-09-974-300-6115	Sequence 6115, Ap
C	29	65	10.4	582	10	US-09-974-300-6100	Sequence 6100, Ap
C	30	65	10.4	954	12	US-10-043-238-2	Sequence 2, Appli
C	31	64.5	10.3	722	10	US-09-925-300-608	Sequence 608, App
C	32	64.5	10.3	1653	10	US-09-815-242-6402	Sequence 6402, Ap
C	33	64.5	10.3	2073	10	US-09-815-242-8362	Sequence 8362, Ap
C	34	64.5	10.3	2109	10	US-09-815-242-4555	Sequence 4555, Ap
C	35	64	10.2	373	10	US-09-783-590-5268	Sequence 5268, Ap
C	36	64	10.2	1341	10	US-09-350-756-2	Sequence 2, Appli
C	37	64	10.2	1740	10	US-09-974-300-479	Sequence 479, App
C	38	64	10.2	4104	10	US-09-801-368-107	Sequence 107, App
C	39	64	10.2	4487	10	US-09-738-968-32	Sequence 32, Appli
C	40	63.5	10.2	670	10	US-09-735-705-7	Sequence 7, Appli
C	41	63.5	10.2	670	10	US-09-850-716A-7	Sequence 7, Appli
C	42	63.5	10.2	670	10	US-09-897-778-7	Sequence 7, Appli
C	43	63.5	10.2	908	12	US-10-044-090-213	Sequence 213, App
C	44	63.5	10.2	1678	10	US-09-925-300-153	Sequence 153, App
C	45	63.5	10.2	1830	10	US-09-815-242-7661	Sequence 7661, Ap

#### ALIGNMENTS

RESULT 1  
US-09-897-438B-1  
; Sequence 1, Application US/09897438B  
; Patent No. US20020137095A1  
; GENERAL INFORMATION:  
; APPLICANT: Mikoshiba, Katsuhiko  
; APPLICANT: Tate, Naoko  
; TITLE OF INVENTION: REELIN PROTEIN CR-50 EPTOPE REGION  
; FILE REFERENCE: 04853-0076-00000  
; CURRENT APPLICATION NUMBER: US/09/897,438B  
; CURRENT FILING DATE: 2001-07-03  
; PRIOR APPLICATION NUMBER: JP 2000-202801  
; PRIOR FILING DATE: 2000-07-04  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 351  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; US-09-897-438B-1

Alignment Scores:  
Pred. No.: 4.49e-79  
Score: 625.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 100.00%  
DB: 10  
Length: 351  
Matches: 117  
Conservative: 0  
Mismatch: 0  
Indels: 0  
Gaps: 0

US-09-897-438B-2 (1-117) X US-09-897-438B-1 (1-351)

Qy	1	GluGlnCysGlyThrIleMetHisGlyAsnAlaValThrPheCysGluProTyrGlyPro	20
Db	1	GAGCAGTGTGGCACCATCATGCAATGGCAATGCTGCACCTTCTGTGAGCGGTACGGCCCT	60
Qy	21	ArgGluLeuThrThrThrCysLeuAsnThrThrThrAlaSerValLeuGlnPheSerIle	40
Db	61	CGAGAGTGACCCACCATGCCCTGAACACAAACAGCATCTCTCTCCAGATTTTCCATT	120
Qy	41	GlySerGlySerCysArgPheSerTyrSerAspProSerIleThrValSerTyrAlaLys	60
Db	121	GGGTGAGGATCATGTGCATTTAGTTACTCTGACCCAGCATCACTGTGTCAACGCCAAG	180
Qy	61	AsnAsnThrAlaAspTrpIleGlnLeuGluLysIleArgAlaProSerAsnValSerThr	80
Db	181	AACAATACCGCTGATTTGGATTACGTGGAGAAAATTAGAGCCCTTCCAAATGTGAGCAC	240
Qy	81	ValIleHisIleLeuTyrLeuProGluGluAlaLysGlyGluSerValGlnPheGlnTrp	100
Db	241	GTCACTCACATCTGTACCTCCCGGAGGAAGCCAAAGGGAGAGCGTGCAGTTCACGTGG	300
Qy	101	LysGlnAspSerLeuArgValGlyGluValTyrGluAlaCysTrpAlaLeu	117
Db	301	AAACAGACACGCTGGCAGTGGGTGAGGTGATGAGCGCTGCTGGGCCCTG	351

RESULT 2  
US-09-880-107-3436  
; Sequence 3436, Application us/09880107  
; Patent No. US20020142981A1  
; GENERAL INFORMATION:  
; APPLICANT: Horne, Darci T.  
; APPLICANT: Vockley, Joseph G.  
; APPLICANT: Scherif, Uwe  
; APPLICANT: Gene Logic, Inc.  
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer  
; FILE REFERENCE: 44921-5028-WO  
; CURRENT APPLICATION NUMBER: US/09/880,107  
; CURRENT FILING DATE: 2001-06-14  
; PRIOR APPLICATION NUMBER: US 60/211,379  
; PRIOR FILING DATE: 2000-06-14  
; PRIOR APPLICATION NUMBER: US 60/237,054  
; PRIOR FILING DATE: 2000-10-02  
; NUMBER OF SEQ ID NOS: 3950  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3436  
; LENGTH: 11580  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 U7977  
US-09-880-107-3436

Alignment Scores:		
Pred. No.:	2.32e-69	11580
Score:	572.00	Matches: 106
Percent Similarity:	95.73%	Conservative: 6
Best Local Similarity:	90.60%	Mismatches: 5
Query Match:	91.52%	Indels: 0
DB:	10	Gaps: 0

US-09-897-438B-2 (1-117) x US-09-880-107-3436 (1-11580)

QY	1	GLuGlnCysGlyThrIleMetHisGlyAsnAlaValThrPheCysGluProTrpGlyPro	20
Db	860	GACAGTGTGGCGGATTATGCATGCAATGCCGTCCACTCTGTGAACCATATGGCCCA	919
QY	21	ArgGluLeuThrThrThrCysLeuAsnThrThrThrAlaSerValLeuGlnPheSerIle	40
Db	920	CGAGAACTGATTACCAACAGCCTTAATACAAACACAGCTCTGTCTCCCAATTTTCCATT	979
QY	41	GlySerGlySerCysArgPheSerTyrSerAspProSerIleThrValSerTyrAlaLys	60

Db	980	GGGTGAGTTTCATGTCGCTTTAGTTATTTCAGACCCGACATCATCGTGTATTATGCCAAG	10399
Qy	61	AsnAsnThrAlaAspTrpIleGlnLeuGluLysIleArgAlaProSerAsnValSerThr	80
Db	1040	AATAACTTCGGGACTGGATTGAGTCAGTAGAGAAAATTAGAGCCCTTCCAATGTCAGCAC	10999
Qy	81	ValIleHisIleLeuTyrLeuProGluGluAlaLysGlyGluSerValGlnPheGlnTrp	100
Db	1100	ATCATCATATCTCTACCTTCCTGAGGAGCGCAAGGGGAGAATGTCCAATTTCACTGG	11599
Qy	101	LysClnAspSerLeuArgValGlyGluValTyrGluAlaCysTrpAlaLeu	117
Db	1160	AACGAGAAAATCTCGTGTAGTGAAGTGTATGAAGCCTGCTGGCCCTT	1210

```

RESULT 3
US-09-864-761-14497/c
; Sequence 14497, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED
; TITLE OF INVENTION: GENE EXPRESSION ANAL
; FILE REFERENCE: Aomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine
; SEQ ID NO 14497
; LENGTH: 499
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC000121.1

```

OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.62  
US-09-864-761-14497

Alignment Scores:  
Pred. No.: 1.52e-22 Length: 499  
Score: 224.00 Matches: 40  
Percent Similarity: 100.00% Conservative: 5  
Best Local Similarity: 88.89% Mismatches: 0  
Query Match: 35.84% Indels: 0  
DB: 10 Gaps: 0

US-09-897-438B-2 (1-117) x US-09-864-761-14497 (1-499)

QY 73 ArgAlaProSerAsnValSerThrValIleHisIleLeuTyrLeuProGluGluAlaLys 92  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 487 AGAGCCCTCCCAATGCAGCAATCATCCATATCTCTACCTCTCGAGGAGCCAA 428  
  
QY 93 GlyGluSerValGlnPheGlnTrpLysGlnAspSerLeuArgValGlyGluValTyrGlu 112  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 427 GGGGAGAATGTCCAATTTCAAGTGAAGCAAAATCTTGTGTAGTGAAGTGTATGAA 368  
  
QY 113 AlaCysTrpAlaLeu 117  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 367 GCCTGCTGGCCCTTA 353

RESULT 4

US-09-864-761-31037/c  
Sequence 31037, Application US/09864761  
Patent No. US20020048763A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharron G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
TITLE REFERENCE: Aeonica-X-1

CURRENT APPLICATION NUMBER: US/09/864,761

CURRENT FILING DATE: 2001-05-23

PRIOR APPLICATION NUMBER: US 60/180,312

PRIOR FILING DATE: 2000-02-04

PRIOR APPLICATION NUMBER: US 60/207,456

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: US 09/632,366

PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: GB 24263.6

PRIOR FILING DATE: 2000-10-04

PRIOR APPLICATION NUMBER: US 60/236,359

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: PCT/US01/00666

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00667

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00664

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00669

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00665

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00668

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00663

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00662

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00670

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: US 60/234,687

PRIOR FILING DATE: 2000-09-21

PRIOR APPLICATION NUMBER: US 09/608,408

PRIOR FILING DATE: 2000-06-30

PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 31037  
LENGTH: 214  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC000121.1  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.62  
OTHER INFORMATION: EST\_HUMAN HIT: BE938667.1, EVALUE 1.00e-76  
OTHER INFORMATION: NT\_HIT: g14826977, EVALUE 1.00e-117  
OTHER INFORMATION: SWISSPROT HIT: Q14063, EVALUE 1.20e+00  
US-09-864-761-31037

Alignment Scores:  
Pred. No.: 2.54e-18 Length: 214  
Score: 190.00 Matches: 33  
Percent Similarity: 100.00% Conservative: 5  
Best Local Similarity: 86.84% Mismatches: 0  
Query Match: 30.40% Indels: 0  
DB: 10 Gaps: 0

US-09-897-438B-2 (1-117) x US-09-864-761-31037 (1-214)

QY 80 ThrValIleHisIleLeuTyrLeuProGluGluAlaLysGlyGluSerValGlnPheGln 99  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 212 ACAATCATCATATCTCTACCTTCTGAGGACGCCAAAGGGAGAGATGTCCAATTTCAG 153  
  
QY 100 TrpLysGlnAspSerLeuArgValGlyGluValTyrGluAlaCysTrpAlaLeu 117  
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 152 TGGACAGCAGGAAATCTCTGTAGGTGATGATGAAGCCTCTGGGCCTTA 99

RESULT 5

US-09-864-761-11808/c

Sequence 11808, Application US/09864761

Patent No. US20020048763A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharron G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO

FILE REFERENCE: Aeonica-X-1

CURRENT APPLICATION NUMBER: US/09/864,761

CURRENT FILING DATE: 2001-05-23

PRIOR APPLICATION NUMBER: US 60/180,312

PRIOR FILING DATE: 2000-02-04

PRIOR APPLICATION NUMBER: US 60/207,456

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: US 09/632,366

PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: GB 24263.6

PRIOR FILING DATE: 2000-10-04

PRIOR APPLICATION NUMBER: US 60/236,359

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: PCT/US01/00666

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00667

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00664

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00669

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00665

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00668

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00663

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00662





; TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
; FILE OF INVENTION: Expression  
; FILE REFERENCE: 10085,500-US  
; CURRENT APPLICATION NUMBER: US/09/974,300  
; CURRENT FILING DATE: 2001-10-05  
; PRIOR APPLICATION NUMBER: 09/680,598  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 60/279,526  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 8481  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2745  
; LENGTH: 795  
; TYPE: DNA  
; ORGANISM: Bacillus licheniformis  
US-09-974-300-2745

Alignment Scores:  
Pred. No.: 0.669 Length: 795  
Score: 72.00 Matches: 22  
Percent Similarity: 41.89% Conservative: 9  
Best Local Similarity: 29.73% Mismatches: 43  
Query Match: 11.52% Indels: 0  
DB: 10 Gaps: 0

US-09-897-438B-2 (1-117) x US-09-974-300-2745 (1-795)

Qy 32 ThrAlaSerValLeuGlnPheSerIleGlySerGlySerCysArgPheSerTyrSerAsp 51  
|||||  
Db 360 ACGGCTTCTAGCGCAATGTCAAAACCGAGCGCGCTTTTCGCTGAGATTAAAG 301  
Qy 52 ProSerIleThrValSerTyrAlaLysAsnThrAlaAspTrpIleGlnLeuGluLys 71  
|||  
Db 300 CCGTATTGGCCGATCCTGCCGCCATGCCAGACAAAGTTGATGATAATGGCCGAAAA 241  
Qy 72 IleArgAlaProSerAsnValSerThrValIleHisIleLeuTyrLeuProGluGluAla 91  
|||  
Db 240 ACAGCGAGACCGATGCCGACGACGAAAAATCAGCCATATTTGATTCGACCCCTTGAAGCA 181  
Qy 92 LysGlyGluSerValGlnPheGlnTrpLysGlnAspSerLeu 105  
|||||  
Db 180 ACAGTCGACTCGATAATCAAGTTGTGGCAAAACGACAGCATATA 139

RESULT 8  
US-09-885-535-3/C  
; Sequence 3, Application US/09885535  
; Patent No. US20020104105A1  
; GENERAL INFORMATION:  
; APPLICANT: Myriad Genetics, Inc.  
; APPLICANT: Heichman, Karen  
; APPLICANT: Bartel, Paul L.  
; TITLE OF INVENTION: Protein-Protein Interactions  
; FILE REFERENCE: 2318-266-II  
; CURRENT APPLICATION NUMBER: US/09/885,535  
; CURRENT FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: US 60/213,245  
; PRIOR FILING DATE: 2000-06-22  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 3  
; LENGTH: 9274  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-885-535-3

Alignment Scores:  
Pred. No.: 50.5 Length: 9274  
Score: 70.00 Matches: 24  
Percent Similarity: 59.09% Conservative: 15  
Best Local Similarity: 36.36% Mismatches: 17  
Query Match: 11.20% Indels: 10  
DB: 10 Gaps: 4

US-09-897-438B-2 (1-117) x US-09-885-535-3 (1-9274)  
Qy 24 ThrThrThrCysLeuAsnThrThrAlaSerValLeuGlnPheSerIleGlySer--- 42  
Db 5024 AGCACATGATGCTTG-----GCAAGTTCAATTTCAATGGCTTTAGGCTCTCCT 4977  
Qy 43 -----GlySerCysArgPheSerTyrSerAspProSerIleThrValSerTyrAlaLys 60  
Db 4976 CCAACAGGTTTCGTC-----TCACTTAGCAAGCCCTCGGTGTGTGTGTCAGCCATGCCAGG 4923  
Qy 61 AsnAsnThr---AlaAspTrpIleGlnLeuGluLysIleArgAlaProSerAsnValSer 79  
Db 4922 AGCTCATCCAGGGCATGTTGGAAGTACGCCAAAGGCTAATAGAGCACCCCTCCAGTTTATGC 4863  
Qy 80 ThrValIleHisIleLeu 85  
Db 4862 TGCTGTGTGATGATTCCTC 4845

#### RESULT 9

US-09-864-761-10582/C  
; Sequence 10582, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Rank, Sharron G.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL F  
; FILE REFERENCE: Aeomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 10582  
; LENGTH: 495  
; TYPE: DNA

US-09-864-761-10582

US-09-897-438B-2 (1-117) x US-09-917-800A-547 (1-468)

Qy            3 CysGlyThrIleMetHisGlyAsnAlaValThrPheCysGluProTyrGlyProArgGlu 22  
               |||            |||      |||      |||||      :::  
Db            393 TGCAGTATTACTACCAAGCGGTGCCTTTACGCCCTGTGAGCCCTGGGG- - - - - 343

Qy	23	LeuThrThrCysLeuAsnThrThr-----ThrAlaSerValLeuGlnPheSerIle	40
Db	342	-----AATGACACACCATTGGACCTCCCTGTGGAGGGCGTTTCAGTCTT	301

QY 41 GlySerGly-----SerCysArgPheSer-----TyrSerAsp 51  
||| ||| ||| ||| |||  
||| ||| ||| ||| ..|||

Ov	52	ProSer-----IleThrValSerTvrA alvAsnAsnThra aaSnTroileGln	68
DB	300	GGATGTGGAAAGGGCTCACCTTCCTGCTCCCGTTCCCTCCTCTCTTGGCAATTTTTTCACCA	24

Db 240 CCCTCCGAGTGGTGCATCAAAAGTCCCCAGCCAGCAACACGGGAGCAAGAGGATTAAG 181

Qy	69	LeuGluLysIleArgAlaProSerAsnValSerThrValIleHisIleLeuTy	88
Db	180	CTAGAGTTTGCATACACGCCCCCAAGCGCCCTTTTGACATCTCTC	136

Qy 89 GluGluAlaLysGlyGlyGluSerValClnPheGlnTrpLysGlnAspSerLeuArgValGly 108

Db	135	GGTGGGGCGGCAGCT-----TGGGATCCCCCTTGCTG---TAGGGG	97
Ov	109	CLNValTurCluAlaCustRoAlaIon	117

Dbb

96 GTTCCCTTTTGTGCCCCCTGGGCTTTG 70

RESULT 11  
US-09-886-055-2  
Sequence 2 Application US/00886055

Patent No. US20020132273A1

APPLICANT: ZOZULYA, SERGEY  
TITLE OF INVENTION: RECEPTOR FINGERPRINTING, SENSORY PERCEPTION, AND  
TITLE OF INVENTION: RECEPTOR FINGERPRINTING, SENSORY PERCEPTION, AND

```

; FILE REFERENCE: 078003-0277150
; CURRENT APPLICATION NUMBER: US/09/886,055
; CURRENT FILING DATE: 2001-06-22

```

; PRIOR APPLICATION NUMBER: 60/213,812  
 ; PRIOR FILING DATE: 2000-06-22  
 ; NUMBER OF SEQ ID NOS: 522

```

; SOFTWARE: PatentIn Ver. 2.1
;
; SEQ ID NO 2
; LENGTH: 079

```

```

; TYPE: DNA
; ORGANISM: Homo sapiens
ttg-00-005-055-3

```

Alignment Scores:

Score:	68.00	Matches:	28
Percent Similarity:	35.25%	Conservative:	21

Query Match:	10.88%	Indels:	48
DB:	10	Gaps:	5

US-09-897-438B-2 (1-117) x US-09-886-055-2 (1-978)

```

QY 2 GlnCysGlyThrIleMethHisGlyAsnAlaValThrPheCysGluProTyrGlyProArg 21
Db 338 AGATGTACTATTTCTATTTGTGTGTGTCATTCAGCAATTTGCTCT---TGGGGACCATGG 394
QY 22 GlnLeuThrThrCysLeuAsnThrThrAlaSerValLeuGlnPheSerIleGly 41
Db 395 CTTATG---ACCATTGTTGGGATCTGCCACCCCTCTGAAATATATACAATTCATGCGGC 451
QY 42 SerGlySer----- 44
Db 452 CCAGGTTGGGCATTTTGCTCACAGTCATCTCATGGTCTCTCAGTAATATTATTGCTCTGA 511
QY 45 -----CysArgPheSerTyrSerAspProSerIleThrValSerTyrAlaLys 60
Db 512 CACACACCCCTTGTGCTCATTTCAATTAATGCTCTTCTGTGAACCACACACACTCTCCACACTTCT 571
QY 61 AsnAsnThrAlaAsp----- 65
Db 572 TCTGTGACTTGGCCCTCTGCTCAAACTGCTCTGTTCAGATCATGTATGATCAATGAGCTTG 631
QY 66 -----TrpIle-----GlnLeuGluLysIleArgAlaPro 75
Db 632 TGTGTTTATTTGGGGTTTATCAGTTATCATCTTCCCTTTACACTCAGCTTCCTTTTCTCT 691
QY 76 SerAsnValSerThrValIleHisLeuTyrLeuProGluGluAlaLysGlyGlu 94
Db 692 ATGCTGTCATCATCAGAGCTGCTCCTCAGAGTATCTTCCACACAGGAAAGTGGAAAG 748

RESULT 12
US-09-802-371-3
; Sequence 3, Application US/09802371
; Patent No. US20010036649A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel
; APPLICANT: Rudolph-Owen, Laura
; TITLE OF INVENTION: 26334, A No. US20010036649A1el Cytidine Deaminase-Like
; FILE REFERENCE: 35800/213921
; CURRENT APPLICATION NUMBER: US/09/802,371
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/188,294
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1017
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-802-371-3

Alignment Scores:
Pred. No.: 4.87 Length: 1017
Score: 67.00 Matches: 25
Percent Similarity: 44.66% Conservative: 21
Best Local Similarity: 24.27% Mismatches: 39
Query Match: 10.72% Indels: 18
DB: 10 Gaps: 4

US-09-897-4388-2 (1-117) x US-09-802-371-3 (1-1017)
QY 2 GlnCysGlyThrIleMethHisGlyAsnAlaValThrPheCysGluProTyrGlyProArg 21
Db 334 CAGATTGCTCTTATTAAACATGGTCAAGGCTCAAAACATGTCATTTATTTTCCAGA 393
QY 22 GlnLeuThrThrCysLeuAsnThrThrAlaSerValLeuGlnPheSerIleGly 41
Db 394 AAACCATGTTCTGCTGT-----TTGAAATGATTTGTAAT 429
QY 42 SerGlySerCysArgPheSerTyr-----SerAspProSerIleThrVal----- 56
Db 430 GCTGGAGTTAACCGAATTCCTACTGCGCTGTCGATGAGAAATGAAGTTTGGTTACGGAG 489
QY 57 ---SerTyrAlaLysAsnAsnThrAlaAspTrpIleGlnLeuGluLysIleArgAlaPro 75

```

## RESULT 14

US-09-974-300-1207/c  
; Sequence 1207, Application US/09974300  
; Patent No. US20020146721A1  
; GENERAL INFORMATION:  
; APPLICANT: Berka, Randy M.  
; APPLICANT: Clausen, Ib Groth  
; TITLE OF INVENTION: Methods for Monitoring Multiple Gene  
; TITLE OF INVENTION: Expression  
; FILE REFERENCE: 10085,500-US  
; CURRENT APPLICATION NUMBER: US/09/974,300  
; CURRENT FILING DATE: 2001-10-05  
; PRIOR APPLICATION NUMBER: 09/680,598  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 60/279,526  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 8481  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1207  
; LENGTH: 612  
; TYPE: DNA  
; ORGANISM: Bacillus licheniformis  
US-09-974-300-1207

## Alignment Scores:

Pred. No.:	2.68	Length:	612
Score:	66.50	Matches:	26
Percent Similarity:	40.59%	Conservative:	15
Best Local Similarity:	25.74%	Mismatches:	44
Query Match:	10.64%	Indels:	16
DB:	10	Gaps:	2

US-09-897-438B-2 (1-117) x US-09-974-300-1207 (1-612)

Qy	3	CysGlyThrIleMetHisGlyAsnAlaValThrPheCysGluProTyrGlyProArgGlu	22
Db	289	TGCTCGCGCTTTAAATGTTCCGGCGCGTATGTCGCTTTCTGT	251
Qy	23	LeuThrThrCysLeuAsnThrThr-----AlaSerValLeu	36
Db	250	-----TCATCCTGTTTTCGGGGTTACCGAACCGTCAATGATTCGGAACCGTTTCG	197
Qy	37	GlnPheSerIleGlySerGlySerCysArgPheSerTyrSerAspProSerIleThrVal	56
Db	196	GAAGAAGCGCAGCTCCGGCCTTGCTCAGCTTATCGTGAGCGTGCGGACATATATCTGTT	137
Qy	57	SerTyrAlaLysAsnAsnThrAlaAspTrpIle-GlnLeuGluLysIleArgAlaProSe	76
Db	136	TCTTCGATGTCAACTCAACCTTGGCGAGCATATCGCCTGCATCCAAATTTTCGACCATG	77
Qy	76	rAsnValSerThrValIleHisIleLeuTyrLeuProGluGluAlaLysGlyGluSerVa	96
Db	76	TACATGATCGTCACGCTGTTTCTCTTTGCTTCCAGATGGCGTAAATGATCGGGGAT	17
Qy	96	1 96	
Db	16	C 16	

## RESULT 15

US-09-741-669-200/c  
; Sequence 200, Application US/09741669  
; Patent No. US2002022718A1  
; GENERAL INFORMATION:  
; APPLICANT: Forsyth, R. Allyn  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; TITLE OF INVENTION: Genes identified as required for  
; TITLE OF INVENTION: proliferation of E. coli  
; FILE REFERENCE: ELITRA.009A  
; CURRENT APPLICATION NUMBER: US/09/741,669  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: US 60/173005  
; PRIOR FILING DATE: 1999-12-23

; NUMBER OF SEQ ID NOS: 481  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 200  
; LENGTH: 2082  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(2082)  
US-09-741-669-200

Alignment Scores:	16.7	Length:	2082
Pred. No.:	66.50	Matches:	25
Score:	44.32%	Conservative:	14
Percent Similarity:	28.41%	Mismatches:	42
Best Local Similarity:	10.64%	Indels:	7
Query Match:	10	Gaps:	4

US-09-897-438B-2 (1-117) x US-09-741-669-200 (1-2082)

Qy	2	GlnCysGlyThrIleMetHisGlyAsnAlaValThrPheCysGluProTyrGly---	Pro 20
Db	1637	AGGTGTAGCTCACCTTGTTAAACGACGCCATCACCGCCTGTTCTCGGACGGTTTCATC	1578
Qy	21	ArgGluLeuThrThrCysLeuAsnThrThrAlaSerValLeuGlnPheSerIle	40
Db	1577	CGCCCGTGTCACGAGCCCAACGTTCAACTCTGTAGCGGCAGTTTCAACTCTTCCAGGTA	1518
Qy	41	GlySerGlySerCysArgPheSerTyrSerAspProSerIleThrVal-----	SerTyr 58
Db	1517	GCTTCCGCGCGCTGCGCTTCCAGCAATTCGACTCTTCAATCAACGTAACAAACCCAGTAT	1458
Qy	59	Ala-----LysAsnAsnThrAlaAspTrpIle-----	GlnLeuGluLysIleArgAla 74
Db	1457	GCCTGACGACCTTCAGTTATGACGCGGTGTCACGCGGTCAATGATGTCGGTACGCGG	1398
Qy	75	ProSerAsnValSerThrValIle	82
Db	1397	GTATCAGGAATAGCGACCGTAGTC	1374

Search completed: November 6, 2002, 22:06:36  
Job time : 91 secs

GenCore version 5.1.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 6, 2002, 19:22:22 ; Search time 2216 seconds  
(without alignments)  
855.087 million cell updates/sec

Title: US-09-897-438b-2

Perfect score: 625

Sequence: 1 EQCGTMRHGNVATFCEPYCP.....FQWKQDSLRLVGEVYEACWAL 117

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_p2n.model -DEV=xlp  
-Q/cgn2\_1/USPTO.spool/US09897438/runat\_06112002\_101924\_11301/app\_query.fasta\_1.263  
-DB=EST -QFMT=fastap -SUFFIX=est -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09897438 @CEN\_1\_1.1716 @runat\_06112002\_101924\_11301 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=3 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	437	69.9	315	12	BE938667	BE938667 QV0-TN008
C 2	401	64.2	600	12	BG803882	BG803882 0243-03 M
C 3	164	26.2	635	10	BB248113	BB248113 BB248113
C 4	132	21.1	568	12	BF387628	BF387628 UI-R-CAL-
C 5	120.5	19.3	1022	17	CNS01272	AL174752 Tetraodon
C 6	96	15.4	979	14	BQ942386	BQ942386 AGENCOURT
C 7	92	14.7	743	10	BB650292	BB650292 BB650292
C 8	87.5	14.0	709	14	BQ444265	BQ444265 UI-M-EXO-
C 9	86.5	13.8	670	10	BB628782	BB628782 BB628782
C 10	86.5	13.8	745	14	BQ180824	BQ180824 UI-M-EXO-
C 11	85	13.6	942	12	BB892165	BB892165 601434685
C 12	84	13.4	369	10	AW158822	AW158822 za45g06.X
C 13	84	13.4	674	13	BI739930	BI739930 603362473
C 14	84	13.4	692	14	BQ769321	BQ769321 UI-M-FIO-
C 15	84	13.4	706	14	BQ771419	BQ771419 UI-M-FIO-
C 16	83	13.3	594	10	BE361240	BE361240 DGL-70-B0
C 17	82.5	13.2	719	14	BQ119839	BQ119839 EST605415
C 18	82.5	13.2	1458	11	AK017094	AK017094 Mus muscu
C 19	81.5	13.0	406	13	BI701324	BI701324 sag57d09.
C 20	81.5	13.0	418	12	BF598079	BF598079 sv03b03.Y
C 21	81.5	13.0	420	13	BI700401	BI700401 sag60d05.
C 22	81.5	13.0	421	14	BQ473635	BQ473635 sap15g06.
C 23	81.5	13.0	449	13	BI700180	BI700180 sag64d10.
C 24	81.5	13.0	529	13	BM307196	BM307196 sak38a07.
C 25	81.5	13.0	546	10	BE440529	BE440529 sp46f03.Y
C 26	80.5	12.9	655	13	BM598586	BM598586 170006876
C 27	80	12.8	455	10	AW036062	AW036062 EST278076
C 28	80	12.8	631	14	BM861182	BM861182 fy46h06.Y
C 29	80	12.8	742	17	AG115800	AG115800 Pan trogl
C 30	79.5	12.7	662	12	BF211911	BF211911 601813083
C 31	79	12.6	1701	13	BM467416	BM467416 AGENCOURT
C 32	78.5	12.6	603	9	AI386086	AI386086 mm17c06.Y
C 33	78.5	12.6	665	10	AV930527	AV930527 AV930527
C 34	78.5	12.6	936	14	BQ882577	BQ882577 AGENCOURT
C 35	78	12.5	511	12	BG378762	BG378762 UI-R-CV1-
C 36	77.5	12.4	477	14	M75904	M75904 CEL06CA Chr
C 37	77.5	12.4	639	17	AZ056749	AZ056749 RPCI-23-4
C 38	77.5	12.4	645	13	BJ501604	BJ501604 BJ501604
C 39	77	12.3	456	9	AA943462	AA943462 EST198961
C 40	77	12.3	539	10	AW141286	AW141286 EST291327
C 41	77	12.3	594	10	AW061669	AW061669 660012H07
C 42	77	12.3	671	17	AG074894	AG074894 Pan trogl
C 43	77	12.3	831	10	BE535985	BE535985 601062413
C 44	76.5	12.2	501	13	BJ220506	BJ220506 BJ220506
C 45	76.5	12.2	723	13	BJ320807	BJ320807 BJ320807

# ALIGNMENTS

RESULT 1  
BE938667/c  
LOCUS BE938667 315 bp mRNA linear EST 02-OCT-2000  
DEFINITION QV0-TN0084-180800-342-a08 TN0084 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BE938667  
VERSION BE938667.1 GI:10466224  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 315)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,

Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare  
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
 Simpson, A.J.  
 Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 20202663  
 Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-QV0-TN0084-180>)  
 800-342-a08t3-2000-08-18t4-1  
 seq primer: puc 18 forward  
 High quality sequence stop: 315.

```

FEATURES
source
1. .315
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="TN0084"
/dev_stage="Adult"
/note="Organ: testis_normal; Vector: puc18; Site_1: Smar;
Site_2: Smar; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
84 a 64 c 78 g 89 t
BASE COUNT
ORIGIN

```

Alignment Scores:				
Pred. No.:	1..43e-45	Length:	315	
Score:	437.00	Matches:	82	
Percent Similarity:	96.67%	Conservative:	5	
Best Local Similarity:	91.11%	Mismatches:	3	
Query Match:	96.92%	Indels:	0	
DB:	12	Gaps:	0	
US-09-897-438B-2 (1-117) x BE938667 (1-315)				
QY	28	LeuAsnThrThrAlaSerValLeuGlnPheSerIleClySerGlySerCysArgPhe	47	
DB	304	CTTAATACACACACACTTCTGCTCTCCAAATTTCCATTTGGGTCAAGTTCATGTCGCTTT	245	
QY	48	SerTyrSerAspProSerIleThrValSerTyrAlaLysAsnAsnThrAlaAspTrpIle	67	
DB	244	AGTTATTACAGCCCAACATCATCGTGTATTATGCCAGAATAACTCTGCGACATGGATT	185	
QY	68	GlnLeuGluLysIleargAlaProSerAsnValSerThrValIleHisIleLeuTyrLeu	87	
DB	184	CAGCTAGAGAAATTAGAGCCCTTCCAATGTCAGACAAATCATCATATCCTCTACCTT	125	
QY	88	ProGluGluAlaLysGlyGluSerValGlnPheGlnTrpLysGlnAspSerLeuArgVal	107	
DB	124	CCTGAGGACGCCAAGGGGAGAATGCCAATTTCACTGGAACGAGGAAGATCTTCGTGTA	65	
QY	108	GlyGluValTyrGluAlaCysTrpAlaLeu	117	
DB	64	GGTGAAGTGTATGAAGCCCTGCTGGGCTTA	35	

RESULT 2	600 bp	mRNA	linear	EST 20-DEC-2001
LOCUS	0243-03	Mouse	retina	lambda ZAP II Library
DEFINITION	musculus cDNA, mRNA sequence.			
ACCESSION	BG803882			

BG803882.1 GI:17950794  
 EST.  
 house mouse.  
 Mus musculus  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.;  
 1 (bases 1 to 500)  
 Mu, X., Zhao, S., Pershad, R., Hsieh, T.-F., Scarpa, A., Wang, S.W.,  
 White, R.A., Beremand, P.D., Thomas, T.L., Gan, L. and Klein, W.H.  
 Gene expression in the developing mouse retina by EST sequencing  
 and microarray analysis  
 Nucleic Acids Res. 29 (24), 4983-4993 (2001)  
 21671825  
 Contact: Klein WH  
 Department of Biochemistry and Molecular Biology  
 University of Texas M.D. Anderson Cancer Center  
 Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA  
 Tel: 713 792 3646  
 Fax: 713 790 0329.

```

tax: 713 790 0329.
FEATURES
  source
    Location/Qualifiers
      1..600
        /organism="Mus musculus"
        /db_xref="taxon:10090"
        /clone_lib="Mouse E14.5 retina lambda ZAP II Library"
        /tissue_type="neural retina"
        /dev_stage="embryonic day 14.5 post-fertilization"
        /note="Vector: pAMP10 (Gibco): Cloned unidirectionally.
Primer: Oligo dT. RNA Isolation: cytoplasmic RNA preps
(Manniatls); Cloning Technique: CUA Cloning (Clontech,
Life Technologies); Average insert size: 1.8 Kb;
Insertion site: TAGCTCAGTGAATCTGAGTG---. Other
information regarding entire library may be found at
http://pga.swmed.edu/Data/Libraries/microarray\_cdna\_library.htm."
      ies.htm.
BASE COUNT      127 a  167 c  152 g  152 t      2 others
ORIGIN

Alignment Scores:
Pred. No.:      1..38e-40      Length:      600
Score:          401.00        Matches:      82
Percent Similarity: 70.09%      Conservative: 0
Best Local Similarity: 70.09%      Mismatches: 3
Query Match:     64.16%      Indels:      33
DB:              12         Gaps:         1

```

```

US-09-897-438B-2 (1-117) x BG6803882 (1-600)

QY      1  GluGlnCysGlyThrIleMethHisGlyAsnAlaValThrPheCysGluProTyrGlyPro  20
       ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DbB     596 GAGCAGTGFGGCACCATCATGATCGCAATGCTGCACCTTCGTGAGCCGTACGGCCCT  537
       ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY      21  ArgGluLeuThrThrThrCysLeuAsnThrThrThrAlaSerValLeuGlnPheSerIle  40
       ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DbB     536 CGAGAGCTGACCACACCATGCTGAAACANCAACAGCATCTGTCTCTCCAGTTTTC---  480
       ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY      41  GlySerGlySerCysArgPheSerTyrSerAspProSerIleThrValSerTyrAlaLys  60
       ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DbB     480 ----- 480

QY      61  AsnAsnThrAlaAspTrpIleGlnLeuGluLysIleArgAlaProSerAsnValSerThr  80
       ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DbB     479 -----ATTGA-GCCCCCTCCAATGTGAGCAC  454

QY      81  ValIleHisIleLeuTyrLeuProGluGluAlaLysGlyGluSerValGlnPheGlnTrp  100
       ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DbB     453 GTTCATCCACATCCTGTACTGCCGAGGAAGCAAGGGAGAGCGTGCAGTTCAGTGG  394
       ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY      101 LysGlnAspSerLeuArgValGlyGluValTyrGluAlaCysTrpAlaLeu  117
       ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DbB     393 AAACAGGACAGCTCGCAGTGGGTGAGGTGATGAGGCCTGCTGGGCCCTG  343
       ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 3
BB248113

```



/strain="Sprague-Dawley"  
 /db\_xref="taxon:10116"  
 /clone="UI-R-CA1-bbs-e-06-0-UI"  
 /clone\_lib="UI-R-CA1"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="Vector: p7T3D-fac (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-CA1  
 library is a subtracted library derived from the following  
 tissues: thalamus, cerebellum, hypothalamus, medulla, pons  
 , midbrain, cerebral cortex, corpus striatum, testis, and  
 hippocampus. For a detailed description of the library  
 from which this clone was derived, please visit our web  
 site at [rateat.eng.uiowa.edu](http://rateat.eng.uiowa.edu). The subtraction has been  
 previously described in (Bonaldo, Lennon and Soares,  
 Genome Research 6:791-806, 1996)

Genome Research 6:791-806, 1996)  
 TAG\_LIB=UI-R-CA1  
 TAG\_TISSUE=pons  
 TAG\_SEQ=AGCAGC"

BASE COUNT 116 a 156 c 131 g 165 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1.84e-06 Length: 568  
 Score: 132.00 Matches: 23  
 Percent Similarity: 96.15% Conservative: 2  
 Best Local Similarity: 88.46% Mismatches: 1  
 Query Match: 21.12% Indels: 0  
 DB: 12 Gaps: 0

US-09-897-438B-2 (1-117) x BF387628 (1-568)

Qy 92 LysGlyGluSerValGlnPheGlnTrpLysGlnAspSerLeuArgValGlyGluValTyr 111

||||| ::|

Db 565 AAGGGAGGAGACGTCACGTTCCAGTGGAAACAGGACACCCCTCGTGGGTGAGGTGAC 506

||||| ::|

Qy 112 GluAlaCysTrpAlaLeu 117

||||| ::|

Db 505 GAAGCCTGCTGGGCCGTG 488

||||| ::|

RESULT 5  
 CNS012TZ  
 LOCUS  
 DEFINITION Tetraodon nigroviridis genome survey sequence PUC-ori end of clone  
 221M15 of library G from Tetraodon nigroviridis, genomic survey  
 sequence.

ACCESSION  
 VERSION AL174752.1 GI:7812809  
 KEYWORDS GSS: genome survey sequence.  
 SOURCE Tetraodon nigroviridis.

ORGANISM  
 Tetraodon nigroviridis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
 Tetraodontidae; Tetraodon.

REFERENCE  
 AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,  
 Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,  
 Saurin,W. and Weissenbach,J.

TITLE  
 Human gene number estimate provided by genome wide analysis using  
 Tetraodon nigroviridis DNA sequence

JOURNAL  
 REFERENCE Unpublished  
 AUTHORS 2 (bases 1 to 1022)

Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,  
 Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and  
 Weissenbach,J.

TITLE  
 Characterization and repeat analysis of the compact genome of the  
 freshwater pufferfish Tetraodon nigroviridis

JOURNAL  
 REFERENCE Unpublished  
 AUTHORS 3 (bases 1 to 1022)

Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-APR-2000)

COMMENT This sequence is a single read and was generated as part of a large

scale clone-end sequencing project of the Tetraodon nigroviridis  
 genome. For more information, please take a look at  
<http://www.genoscope.cns.fr/tetraodon>.

FEATURES  
 source

Location/Qualifiers  
 1..1022  
 /organism="Tetraodon nigroviridis"  
 /db\_xref="taxon:99883"  
 /clone="221M15"  
 /clone\_lib="G"  
 /note="Genoscope sequence ID : COAG221AG08SP1-end ;  
 PUC-ori"

BASE COUNT 246 a 227 c 241 g 307 t 1 others  
 ORIGIN

Alignment Scores:

Pred. No.: 0.000126 Length: 1022  
 Score: 120.50 Matches: 34  
 Percent Similarity: 40.82% Conservative: 6  
 Best Local Similarity: 34.69% Mismatches: 10  
 Query Match: 19.28% Indels: 49  
 DB: 17 Gaps: 3

US-09-897-438B-2 (1-117) x CNS012TZ (1-1022)

Qy 1 GluGlnCysGlyThrIleMethHisGlyAsnAlaValThrPheCysGluProTyrGlyPro 20

||||| ::|

Db 156 GAGCAGTGTGGGTCTCATGCATGGAGGACCGTGACCTCTGTGAACCCCTTGGAGAG 215

||||| ::|

Qy 21 ArgGluLeuThr-----ThrThrCys----- 27

||||| ::|

Db 216 CGAGAGCTGGTAAATCTATGTTGACCGTCGACTTTACATC-TGCACGGAACACCGCCA 274

||||| ::|

Qy 27 ----- 27

Db 275 TCGTCACTGTGACAAAAAGTTGTTCTATGTTTGAATTTGGTTATTAGATTCTTCTG 334

||||| ::|

Qy 28 -----LeuAsnThrThrAlaSerValLeuGlnPheSer 39

||||| ::|

Db 335 TGTGCTGTCTAGGTTACCGTGCCCTCAACACAGCAGCCTCAGTCTCAGTTGCC 394

||||| ::|

Qy 40 IleGlySerGlySer-----CysArgPheSerTyrSer 50

||||| ::|

Db 395 CTCGGTGAAGTTACAGATTATTTTGCAGCATTAATATGCAATCATTTTCT 448

||||| ::|

RESULT 6  
 BQ942386 979 bp mRNA linear EST 21-AUG-2002

LOCUS  
 DEFINITION AGENCOURT\_8800405 NIH\_MGC\_129 Mus musculus cDNA clone IMAGE:6312366

5', mRNA sequence.

ACCESSION  
 VERSION BQ942386.1 GI:22357864

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
 1 (bases 1 to 979)

NTIH-MGC <http://mgs.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

CONTACT: Robert Strausberg, Ph.D.

Email: [cqabps@email.nih.gov](mailto:cqabps@email.nih.gov)

Tissue Procurement: Susan L. Sullivan, PhD.

cDNA Library Preparation: ResGen, Invitrogen Corp

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

plate: LLAMI3737 row: m column: 07

High quality sequence stop: 689.

FEATURES  
 source

Location/Qualifiers  
 1..979



```

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:6312366"
/clone_lib="NIH_MGC_129"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Olfactory epithelium; Vector:
pcmw-SPORT6.1.ccdB; Site_1: EcoRV; Site_2: NotI; Cloned
unidirectionally. Primer: Oligo dt. Average insert size
2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this
is a NIH_MGC Library."
BASE COUNT      232 a      256 c      245 g      245 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:      0.153      Length:      979
Score:          96.00      Matches:    29
Percent Similarity: 47.66%      Conservative: 22
Best Local Similarity: 27.10%      Mismatches: 48
Query Match:    15.36%      Indels:    8
DB:             14      Gaps:      3

US-09-897-438B-2 (1-117) x BQ942386 (1-979)
QY 3 CysGlyThrIleMetHisGlyAsnAlaValThrPheCysGluProTyrGlyProArgGlu 22
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 403 TGTGGTCTGTAGCCAGTGGTAAGCTCTGCTTTTCAACAAGAT---GGGAGGCGCAG 459
QY 23 LeuThrThrCysLeuAsnThrThrAlaSerValLeuGlnPheSerIleGlySer 42
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 460 CTAATCAGTCCTTCTGGACAGCTGCAGCTGCTTTCAGTTTACACTGAGCGTG 519
QY 43 GlySer-----CysArgPhe---SertYrSerAspProSerIleThr 55
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 520 GGGAGCAAGTCTGTGTCAGCAGCTGCAGAGCCCTCACCAGCGGGGGAGTCTGTG 579
QY 56 ValSerTyrAlaLysAsnAsnThrAlaAspTrpIleGlnLeuGluLysIleArgAlaPro 75
: : : : : : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 580 CTGCACATTATCATGACACAGGGGATAACATGGAACCTCTGGAGCACTATTCCTACGTC 639
QY 76 SerAsnValSerThrValIleHisIleLeuTyrIleuProGluGluAlaLysGlyLysSer 95
: : : : : : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 640 AACTACCAGGAGCCAGCAATAATCTCTGTAGAGCTACCGATGATGNCAGACAGTTTTGG 699
QY 96 ValGlnPheGlnTrpLysGln 102
: : : : : : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 700 ATCCAGTTCAGATGCTGCGCAG 720

RESULT 7
BB650292      743 bp      mRNA      linear      EST 26-OCT-2001
LOCUS      BB650292      RIKEN full-length enriched, 0 day neonate cerebellum Mus
DEFINITION      musculus cDNA clone C230087D09 5', mRNA sequence.
ACCESSION      BB650292
VERSION      BB650292.1      GI:16484547
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 743)
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okada
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)

```

```

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-resgsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Imanaka,I., Alzawa
,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences, Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.
FEATURES             Location/Qualifiers
source               1..743
                    /organism="Mus musculus"
                    /db_xref="taxon:10090"
                    /clone="C230087D09"
                    /clone_lib="RIKEN full-length enriched, 0 day neonate
                    cerebellum"
                    /tissue_type="cerebellum"
                    /dev_stage="0 day neonate"
                    /lab_host="DH10B"
                    /note="Site_1: Sall; Site_2: BamHI; cDNA library was
                    prepared and sequenced in Mouse Genome Encyclopedia
                    Project of Genome Exploration Research Group in Riken
                    Genomic Sciences Center and Genome Science Laboratory in
                    RIKEN Division of Experimental Animal Research in Riken
                    contributed to prepare mouse tissues. 1st strand cDNA was
                    primed with a primer [5'
                    GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTNN 3'], cDNA was
                    prepared by using trehalose thermo-activated reverse
                    transcriptase and subsequently enriched for full-length by
                    cap-trapper. cDNA went through one round of normalization
                    to Rot = 20.0 and subtraction to Rot = 479.0. Second
                    strand cDNA was prepared with the primer adapter of
                    sequence [5' GAGAGAGAGATTCGAGTTAATTAATATCCCCCCCCC
                    3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                    modified pBluescript KS(+) after bulk excision from Lambda
                    FLC I."
BASE COUNT      185 a      172 c      225 g      161 t
ORIGIN

Alignment Scores:
Pred. No.:      0.328      Length:      743
Score:          92.00      Matches:    32
Percent Similarity: 43.33%      Conservative: 20
Best Local Similarity: 26.67%      Mismatches: 60
Query Match:    14.72%      Indels:    8
DB:             10      Gaps:      4

US-09-897-438B-2 (1-117) x BB650292 (1-743)
QY 3 CysGlyThrIleMetHisGlyAsnAlaValThrPheCysGluProTyrGlyProArgGlu 22
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 108 TGTGGTCTGTCTTCTGGATCTCTCTGTACTCAGCAAGGCT---GGGAGAGCGCAG 164

```

```

BASE COUNT   159 a   190 c   189 g   170 t   1 others
ORIGIN

Alignment Scores:
Pred. No.:      1.14      Length:      709
Score:          87.50     Matches:     32
Percent Similarity: 40.60%  Conservative: 22
Best Local Similarity: 24.06% Mismatches: 52
Query Match:    14.00%  Indels:    27
DB:             14      Gaps:      5

US-09-897-438B-2 (1-117) x BQ444265 (1-709)

QY 3 CysGlyThrIleMethHisGlyAsnAlaValThrPheCysGluProTyrGlyProArgGlu 22
    |||||::: ||:::||||::: ||| |||
Db 175 TGTGGCTCCACA-----GGCGATGCCCTGGTCTTTATTGMAAAGCCAGCACCCGCTTAC 228
    :::::

QY 23 LeuThrThrThrCysLeuAsnThrThrThrAlaSerValLeuGlnPheSerIleGlySer 42
    :||| ||||| :::: ||| |||||
Db 229 GTGTGCAGCAGACATCGCTGTGAATGAGGACTATCTCTCAGACATAGACTTT---GCT 285
    ||||| :::: ||| |||||

QY 43 GlySerCysArgPheSerTyrSerAspProSerIleThrValSerTyrAlaLysAsnAsn 62
    ||||| :::: ||| |||||
Db 286 GCCTCTCGTCTACGCACAGACTCCTCTGCTATGCTATTCACTGGAGTACTCGGTGGATCTC 345
    :::::

QY 63 ThrAlaAspTrpIleGlnLeuGluLysIleArgAlaProSerAsnValSer---ThrVal 81
    ||| ||| :::: |||:::|||||
Db 346 GGTCTCTCGTGGCACCCTGGTGTGAGGGACTGCCTCCCTACCAATGTTAGTGTAGTCGT 405
    :::::

QY 82 IleHisIle-----Leu 85
    |||:::

Db 406 TACCACCTGCAGCGGATCTGGTGTGCAGATACCTTTCAACAAGTGGACCAAGATCACTCTG 465
    :::::

QY 86 TyrLeuProGluGlnAlaLysGlyGluSerValGlnPheGlnTrp-----100
    ||||| :::: |||:::|||||
Db 466 CCCCTGCCCTCTCTACACCAAGCTCTCAAGGCACCTCGTTTCGGTGGCATCAGCAGCGCCT 525
    :::::

QY 101 -----LysGlnAspSerLeuArgValGlyGluValTyr 111
    ||||| :::: |||||
Db 526 TTTGCAACGACGACGACCTGGGCAATAGATAATGTCTAT 564
    :::::

RESULT 9
BB628782
LOCUS
DEFINITION
ACCESSION
BB628782 670 bp mRNA linear EST 26-Oct-99
BB628782 RIKEN full-length enriched, 16 days neonate cerebellum
musculus cDNA clone 9630032J19 5', mRNA sequence
BB628782

```

VERSION	BB628782.1	GI:16466126
KEYWORDS	EST.	
SOURCE	house mouse	
ORGANISM	mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclerogonathi; Muridae; Murinae; Mus. 1 (bases 1 to 670)	
AUTHORS	Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Muramatsu,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toyota,T., Tsurumai,T. and Hayashizaki,Y.	
TITLE	RIKEN Mouse ESTs (Arakawa,T., et al. 2001)	
JOURNAL	Unpublished (2001)	
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.go.jp, URL:http://genome.qsc.riken.go.jp/	

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)

Wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, Y., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)

Kondo, S., Shingawa, A., Saito, T., Kiyosawa, H., Yamakawa, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. *Mamm. Genome.* 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

e mouse tissues.

FEATURES	SOURCE
1. The first two columns are labeled "FEATURES" and "SOURCE".	
2. The third column is labeled "FEATURES".	
3. The fourth column is labeled "SOURCE".	
4. The fifth column is labeled "FEATURES".	
5. The sixth column is labeled "SOURCE".	
6. The seventh column is labeled "FEATURES".	
7. The eighth column is labeled "SOURCE".	
8. The ninth column is labeled "FEATURES".	
9. The tenth column is labeled "SOURCE".	
10. The eleventh column is labeled "FEATURES".	
11. The twelfth column is labeled "SOURCE".	
12. The thirteenth column is labeled "FEATURES".	
13. The fourteenth column is labeled "SOURCE".	
14. The fifteenth column is labeled "FEATURES".	
15. The sixteenth column is labeled "SOURCE".	
16. The seventeenth column is labeled "FEATURES".	
17. The eighteenth column is labeled "SOURCE".	
18. The nineteenth column is labeled "FEATURES".	
19. The twentieth column is labeled "SOURCE".	
20. The twenty-first column is labeled "FEATURES".	
21. The twenty-second column is labeled "SOURCE".	
22. The twenty-third column is labeled "FEATURES".	
23. The twenty-fourth column is labeled "SOURCE".	
24. The twenty-fifth column is labeled "FEATURES".	
25. The twenty-sixth column is labeled "SOURCE".	
26. The twenty-seventh column is labeled "FEATURES".	
27. The twenty-eighth column is labeled "SOURCE".	
28. The twenty-ninth column is labeled "FEATURES".	
29. The thirtieth column is labeled "SOURCE".	
30. The thirty-first column is labeled "FEATURES".	
31. The thirty-second column is labeled "SOURCE".	
32. The thirty-third column is labeled "FEATURES".	
33. The thirty-fourth column is labeled "SOURCE".	
34. The thirty-fifth column is labeled "FEATURES".	
35. The thirty-sixth column is labeled "SOURCE".	
36. The thirty-seventh column is labeled "FEATURES".	
37. The thirty-eighth column is labeled "SOURCE".	
38. The thirty-ninth column is labeled "FEATURES".	
39. The fortieth column is labeled "SOURCE".	
40. The forty-first column is labeled "FEATURES".	
41. The forty-second column is labeled "SOURCE".	
42. The forty-third column is labeled "FEATURES".	
43. The forty-fourth column is labeled "SOURCE".	
44. The forty-fifth column is labeled "FEATURES".	
45. The forty-sixth column is labeled "SOURCE".	
46. The forty-seventh column is labeled "FEATURES".	
47. The forty-eighth column is labeled "SOURCE".	
48. The forty-ninth column is labeled "FEATURES".	
49. The fiftieth column is labeled "SOURCE".	
50. The fifty-first column is labeled "FEATURES".	
51. The fifty-second column is labeled "SOURCE".	
52. The fifty-third column is labeled "FEATURES".	
53. The fifty-fourth column is labeled "SOURCE".	
54. The fifty-fifth column is labeled "FEATURES".	
55. The fifty-sixth column is labeled "SOURCE".	
56. The fifty-seventh column is labeled "FEATURES".	
57. The fifty-eighth column is labeled "SOURCE".	
58. The fifty-ninth column is labeled "FEATURES".	
59. The sixtieth column is labeled "SOURCE".	
60. The sixty-first column is labeled "FEATURES".	
61. The sixty-second column is labeled "SOURCE".	
62. The sixty-third column is labeled "FEATURES".	
63. The sixty-fourth column is labeled "SOURCE".	
64. The sixty-fifth column is labeled "FEATURES".	
65. The sixty-sixth column is labeled "SOURCE".	
66. The sixty-seventh column is labeled "FEATURES".	
67. The sixty-eighth column is labeled "SOURCE".	
68. The sixty-ninth column is labeled "FEATURES".	
69. The seventieth column is labeled "SOURCE".	
70. The seventy-first column is labeled "FEATURES".	
71. The seventy-second column is labeled "SOURCE".	
72. The seventy-third column is labeled "FEATURES".	
73. The seventy-fourth column is labeled "SOURCE".	
74. The seventy-fifth column is labeled "FEATURES".	
75. The seventy-sixth column is labeled "SOURCE".	
76. The seventy-seventh column is labeled "FEATURES".	
77. The seventy-eighth column is labeled "SOURCE".	
78. The seventy-ninth column is labeled "FEATURES".	
79. The eightieth column is labeled "SOURCE".	
80. The eighty-first column is labeled "FEATURES".	
81. The eighty-second column is labeled "SOURCE".	
82. The eighty-third column is labeled "FEATURES".	
83. The eighty-fourth column is labeled "SOURCE".	
84. The eighty-fifth column is labeled "FEATURES".	
85. The eighty-sixth column is labeled "SOURCE".	
86. The eighty-seventh column is labeled "FEATURES".	
87. The eighty-eighth column is labeled "SOURCE".	
88. The eighty-ninth column is labeled "FEATURES".	
89. The ninetieth column is labeled "SOURCE".	
90. The ninety-first column is labeled "FEATURES".	
91. The ninety-second column is labeled "SOURCE".	
92. The ninety-third column is labeled "FEATURES".	
93. The ninety-fourth column is labeled "SOURCE".	
94. The ninety-fifth column is labeled "FEATURES".	
95. The ninety-sixth column is labeled "SOURCE".	
96. The ninety-seventh column is labeled "FEATURES".	
97. The ninety-eighth column is labeled "SOURCE".	
98. The ninety-ninth column is labeled "FEATURES".	
99. The hundredth column is labeled "SOURCE".	

```

location/quadrifidus
1. .670
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="9630032J19"
/clone_lib="R1EN full-length enriched, 16 days neonate
cerebellum"
/tissue_type="cerebellum"
/dev_stage="16 days neonate"

```

FLC I.	
BASE COUNT	228 a 120 c 114 g 208 t
ORIGIN	

Alignment Scores:		
Pred. No.:	1.41	Length: 670
Score:	86.50	Matches: 29
Percent Similarity:	43.8%	Conservative: 14
Best Local Similarity:	29.5%	Mismatches: 38
Query Match:	13.84%	Indels: 17
DB:	10	Gaps: 4

US-09-897-438B-2 (1-117) x BB628782 (1-670)

Qy	9	GlyAsnAlaValThrPheCysGluProTyrGlyProArgGluLeuThrThrThrCysLeu	28
Db	282	GGAACTTGGAAGACCACTTTGTTTAAGCCCTATTCTTCGGGTCTCGACCCTCACGTGTCTT	341
Qy	29	AsnThrThrAlaSerValLeuGlnPheSerIleGlySerGlySerCysArgPheSer	48
Db	342	CATGTAAAAGTCTCCFCGGGTCTGGTTTGGCATGTGAAAAAGGC---TGTCATGTCGACG	398

Qy	49	TyrSer-----AspProSerIleThrValSerTyrAlaLysAsnAsnThr	63
Db	399	ATGTCACAGACACGAGAGAGACACTCTCTCTCTGAAGTCAGCGCGTTTCTCTAACAAAG	458
		-----GlnLeuLulysIleArgAlaProSerAsnVal	78
Qy	64	AlaAspTrpIle-----	
Db	459	CAGTGCTGGCTAATAGGATACATCTGTAGACTTTGAAATAATTTGGCGCTATTACTCACTCA	518
		-----SerThrValIleHisIleLeuTyrLeuProGluGlu	90
Qy	79	-----	
Db	519	TTCTGAATCGAGAAATAAAGAAATACTACTATTCGCTGTATCTCAAGAACA	572
RESULT	10		
LOCUS	BQ180824		
DEFINITION	745 bp mRNA linear EST 30-APR-2002		
	UI-M-EX0-bxc-m-18-0-UI.r1 NIH_BMAP_EX0 Mus musculus cDNA clone		
	IMAGE:5706425 5', mRNA sequence.		
ACCESSION	BQ180824		
VERSION	BQ180824.1	GI:20356316	
KEYWORDS	EST.		
SOURCE	house musculus		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	1 (bases 1 to 745)		
JOURNAL	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: <a href="mailto:rcgaps@femail.nih.gov">rcgaps@femail.nih.gov</a>		
	Tissue Procurement: Dr. James Lin, University of Iowa		
	cDNA Library preparation: Dr. M. Bento Soares, University of Iowa		
	cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa		
	DNA Sequencing by: Dr. M. Bento Soares, University of Iowa		
	Clone Distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at:		
	<a href="http://image.llnl.gov">http://image.llnl.gov</a>		
	This clone was contributed by the Brain Molecular Anatomy Project		
	(BMAP).		

FEATURES	SOURCE
1. <i>Generalization</i>	1. <i>Generalization</i>
2. <i>Specificity</i>	2. <i>Specificity</i>
3. <i>Consistency</i>	3. <i>Consistency</i>
4. <i>Completeness</i>	4. <i>Completeness</i>
5. <i>Transparency</i>	5. <i>Transparency</i>
6. <i>Interpretability</i>	6. <i>Interpretability</i>
7. <i>Robustness</i>	7. <i>Robustness</i>
8. <i>Scalability</i>	8. <i>Scalability</i>
9. <i>Flexibility</i>	9. <i>Flexibility</i>
10. <i>Adaptability</i>	10. <i>Adaptability</i>
11. <i>Efficiency</i>	11. <i>Efficiency</i>
12. <i>Accuracy</i>	12. <i>Accuracy</i>
13. <i>Reliability</i>	13. <i>Reliability</i>
14. <i>Security</i>	14. <i>Security</i>
15. <i>Privacy</i>	15. <i>Privacy</i>
16. <i>Compliance</i>	16. <i>Compliance</i>
17. <i>Interoperability</i>	17. <i>Interoperability</i>
18. <i>Integration</i>	18. <i>Integration</i>
19. <i>Extensibility</i>	19. <i>Extensibility</i>
20. <i>Customizability</i>	20. <i>Customizability</i>
21. <i>Modularity</i>	21. <i>Modularity</i>
22. <i>Scalability</i>	22. <i>Scalability</i>
23. <i>Flexibility</i>	23. <i>Flexibility</i>
24. <i>Adaptability</i>	24. <i>Adaptability</i>
25. <i>Efficiency</i>	25. <i>Efficiency</i>
26. <i>Accuracy</i>	26. <i>Accuracy</i>
27. <i>Reliability</i>	27. <i>Reliability</i>
28. <i>Security</i>	28. <i>Security</i>
29. <i>Privacy</i>	29. <i>Privacy</i>
30. <i>Compliance</i>	30. <i>Compliance</i>
31. <i>Interoperability</i>	31. <i>Interoperability</i>
32. <i>Integration</i>	32. <i>Integration</i>
33. <i>Extensibility</i>	33. <i>Extensibility</i>
34. <i>Customizability</i>	34. <i>Customizability</i>
35. <i>Modularity</i>	35. <i>Modularity</i>
36. <i>Scalability</i>	36. <i>Scalability</i>
37. <i>Flexibility</i>	37. <i>Flexibility</i>
38. <i>Adaptability</i>	38. <i>Adaptability</i>
39. <i>Efficiency</i>	39. <i>Efficiency</i>
40. <i>Accuracy</i>	40. <i>Accuracy</i>
41. <i>Reliability</i>	41. <i>Reliability</i>
42. <i>Security</i>	42. <i>Security</i>
43. <i>Privacy</i>	43. <i>Privacy</i>
44. <i>Compliance</i>	44. <i>Compliance</i>
45. <i>Interoperability</i>	45. <i>Interoperability</i>
46. <i>Integration</i>	46. <i>Integration</i>
47. <i>Extensibility</i>	47. <i>Extensibility</i>
48. <i>Customizability</i>	48. <i>Customizability</i>
49. <i>Modularity</i>	49. <i>Modularity</i>
50. <i>Scalability</i>	50. <i>Scalability</i>
51. <i>Flexibility</i>	51. <i>Flexibility</i>
52. <i>Adaptability</i>	52. <i>Adaptability</i>
53. <i>Efficiency</i>	53. <i>Efficiency</i>
54. <i>Accuracy</i>	54. <i>Accuracy</i>
55. <i>Reliability</i>	55. <i>Reliability</i>
56. <i>Security</i>	56. <i>Security</i>
57. <i>Privacy</i>	57. <i>Privacy</i>
58. <i>Compliance</i>	58. <i>Compliance</i>
59. <i>Interoperability</i>	59. <i>Interoperability</i>
60. <i>Integration</i>	60. <i>Integration</i>
61. <i>Extensibility</i>	61. <i>Extensibility</i>
62. <i>Customizability</i>	62. <i>Customizability</i>
63. <i>Modularity</i>	63. <i>Modularity</i>
64. <i>Scalability</i>	64. <i>Scalability</i>
65. <i>Flexibility</i>	65. <i>Flexibility</i>
66. <i>Adaptability</i>	66. <i>Adaptability</i>
67. <i>Efficiency</i>	67. <i>Efficiency</i>
68. <i>Accuracy</i>	68. <i>Accuracy</i>
69. <i>Reliability</i>	69. <i>Reliability</i>
70. <i>Security</i>	70. <i>Security</i>
71. <i>Privacy</i>	71. <i>Privacy</i>
72. <i>Compliance</i>	72. <i>Compliance</i>
73. <i>Interoperability</i>	73. <i>Interoperability</i>
74. <i>Integration</i>	74. <i>Integration</i>
75. <i>Extensibility</i>	75. <i>Extensibility</i>
76. <i>Customizability</i>	76. <i>Customizability</i>
77. <i>Modularity</i>	77. <i>Modularity</i>
78. <i>Scalability</i>	78. <i>Scalability</i>
79. <i>Flexibility</i>	79. <i>Flexibility</i>
80. <i>Adaptability</i>	80. <i>Adaptability</i>
81. <i>Efficiency</i>	81. <i>Efficiency</i>
82. <i>Accuracy</i>	82. <i>Accuracy</i>
83. <i>Reliability</i>	83. <i>Reliability</i>
84. <i>Security</i>	84. <i>Security</i>
85. <i>Privacy</i>	85. <i>Privacy</i>
86. <i>Compliance</i>	86. <i>Compliance</i>
87. <i>Interoperability</i>	87. <i>Interoperability</i>
88. <i>Integration</i>	88. <i>Integration</i>
89. <i>Extensibility</i>	89. <i>Extensibility</i>
90. <i>Customizability</i>	90. <i>Customizability</i>
91. <i>Modularity</i>	91. <i>Modularity</i>
92. <i>Scalability</i>	92. <i>Scalability</i>
93. <i>Flexibility</i>	93. <i>Flexibility</i>
94. <i>Adaptability</i>	94. <i>Adaptability</i>
95. <i>Efficiency</i>	95. <i>Efficiency</i>
96. <i>Accuracy</i>	96. <i>Accuracy</i>
97. <i>Reliability</i>	97. <i>Reliability</i>
98. <i>Security</i>	98. <i>Security</i>
99. <i>Privacy</i>	99. <i>Privacy</i>
100. <i>Compliance</i>	100. <i>Compliance</i>

```

1. 1743
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:5706425"
/clone.lib="NIH_BMAP_EX0"
/tissue_type="whole brain"
/dev_stage="embryo 15.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/note="Organ: Brain; Vector: pVX-Asc; site_1: EcoR I;
site_2: Not I; The library was constructed according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured mRNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was size selected according to mRNA size fraction,
ligated with EcoR I adaptor, digested with Not I, and then
cloned directionally into pVX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is GTCCGCGGAA. This library was created for the
University of Iowa Mouse Brain Molecular Anatomy Project
(BMAP): 'Gene Discovery in the Developing Mouse Nervous
System', supported by National Institutes of Mental Health
(NIMH), Hemin Chin, Ph.D., program coordinator."
199 c 207 q 178 t

```

BASE COUNT	159 a	199 c	207 g	178 t	2 o
ORIGIN					
Alignment Scores:					
Pred. No.:	1.64				Length:
Score:	86.50				Matches:
Percent Similarity:	40.60%				Conservative:
Best Local Similarity:	24.06%				Mismatches:

```
Query Match: 13.84% Indels: 27
DB: 14 Gaps: 5
US-09-897-438B-2 (1-117) x B0180824 (1-745)
QY 3 CysGlyThrIleMethHisGlyAsnAlaValThrPheCysGluProTyrGlyProArgGlu 22
    |||||::: ||:::||||::: ||| |||
DB 85 TGTGGCTCCACA-----GGCGATGCCCTGGTCTTTATTGAAAGGCCAGCACCGCTTAC 138
QY 23 LeuThrThrCysLeuAsnThrThrAlaSerValLeuGlnPheSerIleGlySer 42
    :::: |||||::: ||| ||||| |||||
DB 139 GTGGTCACGACAGACCTCCTGTGAATGAGGACTCATCTTCACAGATAGACTTTT---GCT 195
QY 43 GlySerCysArgPheSerTyrSerAspProSerIleThrValSerTyrAlaLysAsnAsn 62
    |||||::: ||| |||||::: ||| |||||::: |||
DB 196 GCCTCCTGCTAGTCACACACTCCCTGCTATGCTATTGAACCTGGACTCGGTGGATCTC 255
QY 63 ThrAlaAspTrpIleGlnLeuGluLysIleArgAlaProSerAsnValSer---ThrVal 81
    ||| ||| ||| ||| ||||| |||||
DB 256 GGTCTGTCTGGCACCGCTGTGTGAGGACTGCCTGCCTACCAATGTTGAGTGTAGTCGT 315
QY 82 IleHisIle-----SerIleGlySer-----GlySerCysArgPheSerTyr 49
    |||||::: ||||| ||||| ||||| |||||
DB 316 TACCACCTGCACGGGATCCTGTGTACAGATACTTTCAACAAGTGGACCAAGATCACTCTG 375
QY 86 TyrLeuProGluGluAlaLysGlyGluSerValGlnPheGlnTrp----- 100
    |||||::: ||| |||||::: ||||| |||||
DB 376 CCCTGCTCTTCCACACAGCTCTCAAGCCACTGTTCCGCTGGCATCGACCGCGCT 435
QY 101 -----LysGlnAspSerLeuArgValGlyGluValTyr 111
    |||||::: ||| ||||| |||||
DB 436 TTTGACAGCAGCAGACCTGGGCAATAGATNATGCTAT 474
RESULT 11
BE892165/c
LOCUS BE892165 942 bp mRNA linear EST 20-OCT-2000
DEFINITION 601434685F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3919964 5',
mRNA sequence.
ACCESSION BE892165
VERSION BE892165.1 GI:10352223
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 942)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC/DCTD/DTF
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM9750 row: i column: 21
High quality sequence stop: 577.
Location/Qualifiers
1. .942
/organism="Homo sapiens"
/db.xref="taxon:9606"
/clone="IMAGE:3919964"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2 kb. Library constructed by Life
Technologies."
BASE COUNT 260 a 258 c 234 g 190 t
```

```
ORIGIN
Alignment Scores: 3.61 Length: 942
Pred. No.: 85.00 Matches: 33
Score: 33.33% Conservative: 19
Percent Similarity: 21.15% Mismatches: 52
Best Local Similarity: 13.60% Indels: 52
Query Match: 12 Gaps: 8
DB: 12
US-09-897-438B-2 (1-117) x BE892165 (1-942)
QY 3 CysGlyThr-----IleMethHisGlyAsnAlaValThrPheCysGluProTyrGlyPro 20
    |||||::: ||:::||||::: ||| |||||
DB 838 TGTGGCCCGCCGGCGTGTGCATAGTTCCTCGGTG---CTGTTCGTCGGCGCGCCTT 782
QY 21 ArgGluLeuThrThrCysLeuAsnThrThrAlaSerValLeuGlnPhe----- 38
    ||| ||| ||| ||| ||||| |||||
DB 781 CGTTCCTGTGCAGCTGCGTGTCTTTTTCGCCCTGTGCGCATGTTTGACATACCCGTA 722
QY 39 -----SerIleGlySer-----GlySerCysArgPheSerTyr 49
    ||||| ||||| ||||| ||||| |||||
DB 721 CGTCCGCGCAGCGTCTCTTCGATTGTAATTTCTTCAGGGGATCCTGTAGACCTGCTCTA 662
QY 50 SerAspProSerIleThrValSerTyrAlaLysAsn----- 61
    |||||::: ||| |||||::: ||||| |||||
DB 661 TGAACCCCTTTAGTAACATTAGATATCTTGCAAGCAGATATTCTTATGGATAAGGCAC 602
QY 62 -----AsnThrAla 64
    |||||::: ||| |||||::: ||||| |||||
DB 601 AGCTTCCAGGATGCTGCGCTCTGGATGTTTCAGAGGAGGAGAAAGGTGTCAAAACAGCG 542
QY 65 AspTrpIleGlnLeuGluLys-----IleArgAlaProSerAsnValSerThr 80
    ||||| ||||| ||||| ||||| |||||
DB 541 GACTGGTTAAATTTTCCCAAAACTGTTGTTGGTTCGAGGACCAAGCA----- 494
QY 81 ValIleHisIleLeuTyrLeuProGluAlaLysGlyGluSerValGlnPheGlnTrp 100
    :::: ||| ||| ||| ||| |||||
DB 493 -----TGGCACTGTCCTCCATCCATCTGCCACAGAGCATGCAAAAGGCAATGG 446
QY 101 LysGlnAspSerLeuArgValGlyGluValTyrGluAla---CysTrp 115
    ::: |||||::: ||||| ||||| |||||
DB 445 CGGACGATGACAGAGGGGTGGCCACCAAGTATCATCTGCTGTGG 398
RESULT 12
AW158822
LOCUS AW158822 369 bp mRNA linear EST 05-NOV-1999
DEFINITION za45g06.x1 Xenopus EST library Xenopus laevis cDNA clone za45g06
5', mRNA sequence.
ACCESSION AW158822
VERSION AW158822.1 GI:6270851
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 369)
AUTHORS Schutz,K., de la Bastide,M., Huang,E.N., Nascimento,L., Preston,R.,
Shah,R., Swaby,I., Shekher,M., Spiegel,L., Vil,M.D. and McCombie
,W.R.
Expressed sequence tags from Xenopus
Unpublished (1999)
JOURNAL
COMMENT Contact: W. Richard McCombie
Lita Annenberg Hazen Genome Sequencing Center
Cold Spring Harbor Laboratory
PO Box 100, Cold Spring Harbor, NY 11724, USA
Tel: 516 367 8884
Fax: 516 367 8874
Email: mcombie@cshl.org
Plate: za45 row: g column: 06
Seq primer: M13 universal forward primer
High quality sequence stop: 369.
```

Tissue Procurement: The Cepko Laboratory  
cDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM11940 row: a column: 20  
High quality sequence stop: 635.

FEATURES  
Location/Qualifiers  
1..674  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5369371"  
/clone\_lib="NIH\_MGC\_94"  
/tissue\_type="retina"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: eye; Vector: pCMV-SPORT6; Site\_1: Noti;  
Site\_2: Salt; Cloned unidirectionally; oligo-dT primed  
Average insert size 3.3 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC library."

BASE COUNT 150 a 177 c 189 g 158 t  
ORIGIN

Alignment Scores:

Pred. No.:	2.94	Length:	674
Score:	84.00	Matches:	34
Percent Similarity:	43.55%	Conservative:	20
Best Local Similarity:	27.42%	Mismatches:	50
Query Match:	13.44%	Indels:	20
DB:	13	Gaps:	8

US-09-897-438B-2 (1-117) x BI739930 (1-674)

```

Qy   3  CysGlyThrIleMethHisGlyAsnAlaValThrPhe-----CysGluProTyrgly 19
      ||||| ::: ||| |||::: ||| |||
Db   194 TGTGGGCTGTGGCTTCGGCGCTCCAGTTCATTCAATGGGGCTGCAGC----- 244

Qy   20 ProArgLeuLeuThrThrThrCysLeuAsnThrThrThralaserValLeuGlnPheSer 39
      |||:: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   245 --CGATGTTGATCACTGGATCTGGAACTCAACCTCACCAATGCTGAGTTATCCAGTTTTAC 301

Qy   40 IleGlySerGlySerCysArgPheSerTyrSer-----AspProSerIleThrValSer 57
      |||:: ||| ||| ::: ||| ::: ||| ::: |||
Db   302 TTATGTTATGA--TGCCCTCATTCGCCGCGCAACCCTAACCCAGGGAGTCTGCTGGAG 358

Qy   58 TyrAlaLysAsnAsnThrAlaAspTrp---IleGlnLeuGluLysIleArgAlaProSer 76
      |||::: ||| ||| ::: ||| ::: ||| ::: ||| ||| ||| ||| ||| |||
Db   359 TACTCTGCAATGAGGAGCATCACTCGAAGCTTGCTGATGGAGATGTTCTATGACCAGTAC 418

Qy   77 AsnValSerThrValIleHisIleLeuTyrlLeuProGluGluAlalysGlyGluSerVal 96
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   419 AGCAAACCTCGATTTGTGAATACTCCCT---CTCCCCTCTGATGCTAAAGAGATGGCACT 475

Qy   97 GlnPheGlnTrp-----LysGlnAspSerLeuArgValGlyGluValTyrGluAla 113
      ::|||:::||| ||| ::: ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   476 CGCTTCGATGGTGGCAGCCACGACATGATGGCCTTGACCAGATGAC----- 523

Qy   114 CysTrpAlaLeu 117
       |||||:::
Db   524 ---TGGGCCATT 532

RESULT 14
BQ769321
LOCUS
DEFINITION
ACCESSION BQ769321
VERSION BQ769321.1 GI:21977795
KEYWORDS EST.
SOURCE house mouse
ORGANISM Mus musculus
BQ769321
UT-M-F10-by-p-n-15-0-UI.r1 NIH_BMAP_F10 Mus musculus cDNA clone
IMAGE:5721422 5', mRNA sequence.

```





**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 13:07:41 ; Search time 2762 Seconds  
(without alignments)  
3698.438 Million cell updates/sec

Title: US-09-897-438B-1  
Perfect score: 351  
Sequence: 1 gagcagtgtggcaccatcat.....atgaggcctgtgggcccctg 351

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.in.\*
- 18: em.hum.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.in.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pln.\*
- 35: em.htg.rod.\*
- 36: em.htg.mam.\*
- 37: em.htg.vrt.\*
- 38: em.sy.\*
- 39: em.htgo.hum.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	351	100.0	11673	6	AX305377	AX305377 Sequence
2	351	100.0	11673	10	MMU24703	U24703 Mus musculus
3	299.8	85.4	11187	10	AB049473	AB049473 Rattus no
4	263.6	75.1	11580	6	AX410790	AX410790 Sequence
5	263.6	75.1	11580	9	HSU79716	U79716 Human reeli
6	263.6	75.1	11580	11	G30936	G30936 SWS2926 Er
7	263.6	75.1	11580	11	G30938	G30938 SWS3176 Er
8	189.4	54.0	10634	5	AF090441	AF090441 Gallus ga
9	135	38.5	163985	10	AC121878	AC121878 Mus muscu
10	135	38.5	163985	2	AC023062	AC023062 Mus muscu
11	115.8	33.0	136130	2	AC095877	AC095877 Rattus no
12	115.8	33.0	202764	2	AC128022	AC128022 Rattus no
13	114.2	32.5	183641	2	AC124933	AC124933 Rattus no
14	98.8	28.1	93163	9	HSAC000121	AC000121 Human BAC
15	53.8	15.3	185996	2	AC041023	AC041023 Homo sapi
16	52.2	14.9	158	4	AF232904	AF232904 Bos tauru
17	37.2	10.6	2662	5	GGA317960	AJ317960 Gallus ga
18	37.2	10.6	145264	9	AC107939	AC107939 Homo sapi
19	36.6	10.4	90166	2	AC114052	AC114052 Rattus no
20	36.6	10.4	143517	9	AL356752	AL356752 Human DNA
21	36.6	10.4	195670	2	AC115449	AC115449 Rattus no
22	36.4	10.4	55452	3	AC004328	AC004328 Drosophil
23	36.4	10.4	63751	2	AC017489	AC017489 Drosophil
24	36.4	10.4	164567	3	AC091500	AC091500 Drosophil
25	36.4	10.4	246095	3	AE003802	AE003802 Drosophil
26	35.8	10.2	173841	10	AL732391	AL732391 Mouse DNA
27	35.8	10.2	189541	10	AC093447	AC093447 Mus muscu
28	35.6	10.1	1365	9	HSJ308525	AJ308525 Homo sapi
29	35.6	10.1	1812	10	AF245444	AF245444 Mus muscu
30	35.6	10.1	2619	9	HSJ299451	AJ299451 Homo sapi
31	35.6	10.1	2973	9	HSJ249210	AJ249210 Homo sapi
32	35.6	10.1	3200	10	RATGLUR7A	M83552 Rattus norv
33	35.6	10.1	3344	6	AR001481	AR001481 Sequence
34	35.6	10.1	3380	11	G26866	G26866 human STS S
35	35.6	10.1	3593	10	RNGLUR7	Z11716 R. norvegicu
36	35.6	10.1	3614	9	HSU16127	U16127 Human gluta
37	35.6	10.1	3619	6	AR180924	AR180924 Sequence
38	35.6	10.1	3633	10	AF027331	AF027331 Rattus no
39	35.2	10.0	14089	10	AF060195	AF060195 Mus muscu
40	35.2	10.0	34849	10	AB053120	AB053120 Mus muscu
41	35.2	10.0	148846	10	AL591204	AL591204 Mouse DNA
42	35	10.0	87370	2	AL356412	AL356412 Homo sapi
43	35	10.0	153056	10	AL607090	AL607090 Mouse DNA
44	35	10.0	160557	9	AC117945	AC117945 Homo sapi
45	35	10.0	193468	2	AC125591	AC125591 Rattus no

ALIGNMENTS

RESULT 1	AX305377	Sequence 128 from Patent WO0188188.	11673 bp	DNA	linear	PAT 11-DEC-2001
LOCUS	AX305377	Sequence 128 from Patent WO0188188.				
DEFINITION	AX305377					
ACCESSION	AX305377					
VERSION	AX305377.1	GI:17644926				
KEYWORDS						
SOURCE	house mouse.					
ORGANISM	Mus musculus					
REFERENCE	1					
AUTHORS	Ishikawa, K., Asai, S., Takahashi, Y., Nagata, T. and Ishii, Y.					
TITLE	Method for examining ischemic conditions					
JOURNAL	Patent: WO 0188188-A 128 22-NOV-2001;					

School Juridical Person Nihon University (JP)

FEATURES

source  
1. .11673  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
BASE COUNT 2831 a 2985 c 2985 g 2872 t  
ORIGIN

Query Match 100.0%; Score 351; DB 6; Length 11673;  
Best Local Similarity 100.0%; Pred. No. 5.7e-96;  
Matches 351; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAGCAGTGGCAGCATCATGCGAATCTGTGACCTTCTGTGAGCCGTACGGCCCT 60  
Db 970 GAGCAGTGGCAGCATCATGCGAATCTGTGACCTTCTGTGAGCCGTACGGCCCT 1029  
Qy 61 CGAGAGCTGACACACATGCTGAACACACACACATGCTGCTCCAGTTTCCATT 120  
Db 1030 CGAGAGCTGACACACATGCTGAACACACACACATGCTGCTCCAGTTTCCATT 1089  
Qy 121 GGGTCAGGATCATGTCGATTAGTTACTCTGACCCAGCATCACTGTGTACAGCCCAAG 180  
Db 1090 GGGTCAGGATCATGTCGATTAGTTACTCTGACCCAGCATCACTGTGTACAGCCCAAG 1149  
Qy 181 AACATACCGCTGATTGATTACCTGAGGAAATAGAGCCCTTCCATGTGAGCACA 240  
Db 1150 AACATACCGCTGATTGATTACCTGAGGAAATAGAGCCCTTCCATGTGAGCACA 1209  
Qy 241 GTCATCCATCCTGATGAGTGGGTGAGTGTATGAGCCCTGCTGGGCCCT 351  
Db 1210 GTCATCCATCCTGATGAGTGGGTGAGTGTATGAGCCCTGCTGGGCCCT 1269  
Qy 301 AAACAGCAGACCCGCGAGTGGGTGAGTGTATGAGCCCTGCTGGGCCCT 351  
Db 1270 AAACAGCAGACCCGCGAGTGGGTGAGTGTATGAGCCCTGCTGGGCCCT 1320

RESULT 2

MMU24703 11673 bp mRNA linear ROD 19-DEC-1997  
DEFINITION Mus musculus reelin mRNA, complete cds.  
ACCESSION U24703  
VERSION U24703.1 GI:2702252  
KEYWORDS  
SOURCE mouse.

ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 11673)  
D'Arcangelo, G., Miao, G.G., Chen, S.C., Soares, H.D., Morgan, J.I. and Curran, T.  
A protein related to extracellular matrix proteins deleted in the mouse mutant reeler.  
Nature 374 (6524), 719-723 (1995)  
95231649  
7715726

REFERENCE

1 (bases 1 to 11673)  
D'Arcangelo, G.  
Direct Submission  
Submitted (10-APR-1995) Roche Institute of Molecular Biology, 340 Kingsland St., Nutley, NJ 07110, USA  
3 (bases 1 to 11673)  
D'Arcangelo, G.  
Direct Submission  
Submitted (19-DEC-1997) Dev. Neurobiology, St. Jude Childr. Res. Hosp., 332 N. Lauderdale, Memphis, TN 38105, USA

AUTHORS

Sequence update by submitter  
On Dec 19, 1997 this sequence version replaced gi:902486.  
Location/Qualifiers  
1. .11673  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="5"

REMARK

COMMENT  
On Dec 19, 1997 this sequence version replaced gi:902486.  
FEATURES

source

1. .11673  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="5"

3'UTR  
polyA\_signal  
BASE COUNT 2831 a 2985 c 2985 g 2872 t  
ORIGIN

/map="D5Gmr1"  
/tissue\_type="cerebellum"  
/dev\_stage="adult"  
1. 282  
283. .10668  
/codon\_start=1  
/product="reelin"  
/protein\_id="AAB91599.1"  
/db\_xref="GI:2702253"  
/translation="MERGOWAPRALVLAIVLLIATLARAATGYPRFPFFFLCTHH  
GELEGGEGEVLISHIAGNPYYPGGEYHVHTIISTFTFDGLLVLTSTISQSS  
QSTGGSAFQIMSDHQFQMCVSAVSHVSHLPTNLFSVIMAPPAGTCGVNPM  
TATPHRGQVIFKDALAQOECQAPATAYSHLAEIHSDFVILRDDPSYQOLEP  
IWECSNCEMGECGTTIMHGNATFCEPRLTTLTCLNTTASVQLQSFSGSCRF  
SYSDPTISVYACQNTADMIOLKIRAPSNVSTVILYLPPEAKESVQFQKQDSL  
RGEVVEACWALNDILVINSAREVLEONLDPDGTGNMLFFPGATVKHSCSDGNSI  
YFNGEGSEFNATTARDVDLSTEDIEIQBSEFEESQPTGWDILGAVGADCGTESGL  
SIVFLDKGERKKCTPYMDTQYGNLRFYFVMSGICDPGVSHENDIILYAIKGRKEHI  
ALDTLYSSYKPSLVSVINPELQTPATKFNLRKSHOGYNRNVAWDFHVLPTLP  
STMSHMIQFSINLGGCTHQFNSVLEFSTNKLGRKWSLLHTECLPEICAGPLHPTV  
YSSENTSGWNRITIPLPNAALTDRTRIWRQTGPILGNMAIDNVYISGAEVFCGSR  
GQCTRHGCKDCPGFCACEMASQTPMFISESFGSARLSSVHNFSIRGAEVFCGG  
VLASGKALVFNKDRRLQTLTSLDSSQSRFLQTLRLGSKSVLSTCRADPQEGEVL  
HYSIDNGITWKLHYSYVNHPEIISVELPDDARQFQIFRWMPQYHSSQEDWMA  
IDIVMTSVLNSISLDTNLVEVTSGLFYLGNOVQYCGHDTWLTCTGDSKLASSM  
YVETQSMQIGASYMIQFSVLMGCGQYTPHMDNQVKLEYSAHSHYTLHWLQEBCLPSM  
PSCQETASISYHASEFTOWRRVTVLPKOTWSGATFRWSQSYTYAQDEWALNDIYI  
GQCPNMGSGHSCDGHGCRDOGYGTCHPEALPSTIMSDFENPSSWESQOEVI  
GEBVPEOGCGVVSQSSLYFSKAGKRLQVSHDLTSDWDFVQFYIQGGESAACNK  
PSREGEILQYNSNGGIQWHLAEMFSDFSKPREVYIELPAAGTKPTCFRFRWPKP  
FSGEDYQWAVDDIILSEKQOVIPIVNPPTLPQNFYKPAEDFPMNQMSVWMLANE  
GNAKNDSCATTSPAMVFGKSDGDFRATDLTKPGYVLQKINCGTCSQFSTAPV  
LIQSHDAGMSWFLKEGCPASAAKCGENRELSPTVYVYTGDEEWTITAIAPR  
SLASKTRFRWFLQESSOKNVPFGLDGVYISPCPSYCGHGDCTSGVCFCDLGYTA  
AGCTCVSNTNHESEMEDREGKLSPLWYKITGGVCTGCTLNDGRSLFENGLKREA  
RIVPLDTRNISLVQFYIQIGSKTSGITYITPRARYEGLVYVSGGVTHWLMDEFY  
MSFLEPIIISIDLPREAKTATAFRMOQHGKSAOWALGDVLIGNDSSQTFQDK  
LDGSDIQANWYRIGQGVDDICLSMDTALITENIGNPYRATWDFHVSSEFLQWE  
MNGCSKPFSGAHQIOQLSNNGKDWLVTECVPTTGVHTYESSVTSERFQNW  
RVTVYPLATNSPRTFRWITNTVAGADWADNVILASGCPWMSGRCIGCDGRC  
LYDVRGPGGCPVVPVPLPSILKDDFNGNLHPOLPEVYGAERGNLNGETIKSGTCLIF  
KGEGLMLISRLDCTNTWYVQSLRFIAKTPERSHSILQFVSQGVTHWLMDEFY  
FQPTTSILFINVPLPYGAQTNATFRMLQPYNNGKKEIWIIDDFIIGNNUNNPVLL  
LDTDFEPREDNFFYPGGNIGLYCPYSKGAPEEDSAMVFSVEGHSITTDLSV  
NENTIIQFEINVCSDSDSADPVRFSRDFCATWHLLPLCYHSSSLVSLCSTE  
HPSTVYAGTQGWREVYFGKLHCGSVFRWYOGFYAGSOPVTWADNVYIGPQ  
CEEMCYGHSCINGKICIDPGYSGTCKISTKNPDFLKDDEGQLESREFLMSGGK  
PBRKCGILSGNNLFFNEDGLRMLVTRDLDLSHAREVQFMRIGCGCKGVDPDRSQVPL  
LOYSNLGSLWSLLQBFLEFSNSNVGRYIALEMLKARSGSTRLRMWQPSRNGHYSP  
YVVTDDIANDSDSFLIDFAASCSTVDSCYATELYSVDLGLSWHPLVRDCLPTNVEC  
SRVHLQRIIVSDTFNKWTRITILPLPSYTRSOATFRFHWOPAPDKOQWADNVYIGD  
GCLDMSGHCRCVQSCGDEQWGLLYCDEPETSILPTQKDNFNAPSNQMLVTSVG  
KLSTVCGAVASGLAHLHSGGCSRLVTVDLNLTNAEYIQFYFMNGLIPLPSNRNQVL  
LEYSYNGITWNLMEIFYDYSKPGFNILLPPDAKETAIRFRMWQPRHGLDQNDW  
AIDNVILSGSADQRTVMDTFSAPVQPHRSPADAPGVGATFAMFLBDKTSVNECW  
LFHDDCTVERFCDSPGVMLCGHDGREVYAVTHDITPTENIMOFKISVGCCKVPEKI  
AQONITHQVSTDFGYSWYLVQCLPADPKCSGSYSQSPYFETEGWKKRITPLPSL  
TGNVRFREYQKISDVQWADNVYLGPCDNCGGGCKECCICDPIYSGPNCTYLT  
HSLKTLKERFDSBEIKPDLMSLEGGSTCEGVAENTALYFGSTVYRQATQDLD  
LRGAKFLQYWRIGRISNNMTSCHRPVREGVLLDFTSDGTITWTLHMDQKTLISV  
RHDYTLPEGALNTTRLRWMPQFVINGVLSVGRVQAWALNDILIGAEINPSOLV  
DTFDDGSSHEENWSPYNAVRTAGFCGNPSFHLWPKKDKKTINALSSRELITQPG  
YMKQPIVYCCATSCGDLHSHVMEYTKDARSQVQLVQVQCLPSSNSIGCSPPQFH  
EATITNAVNSWKKRITIQLPDHSVSSATQFRWIKQETEKOSWAIDNVYIGEACP  
LCSGHGYCTTGAVCDESDQDCSVFSHELPSYIKDNFESARTANWETIQQGCPK  
SGCGGLAPYAHGDSLYNGCQIRQAATKPLDLTRASKIMFYLGTSQPSQSCNDL  
SGCPHTVDRKAVLLQYSVNNGTIWHVTAQHKPKDQTOARVSNVPLEARKMGLVLRMWQ  
PRNGTGHQWALDHVEVVLVSTRKQNYMWNFSRQHLRHFYNNRRRSRLRPY"  
10669. .11673  
11662. .11667

Query Match 100.0%; Score 351; DB 10; Length 11673;  
Best Local Similarity 100.0%; Pred. No. 5.7e-96;  
Matches 351; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGCAGTGTGGCACCACCATCATGCGAATGCTGTACCTTCTGTGAGCGCGTACGGCCCT 60  
|||||  
Db 970 GAGCAGTGTGGCACCACCATCATGCGAATGCTGTACCTTCTGTGAGCGCGTACGGCCCT 1029  
|||||

QY 61 CGAGAGCTGACACACACATGCTGAAACACAAACAGCAGCATCTCTCCATGTTTCCATT 120  
|||||  
Db 1030 CGAGAGCTGACACACACATGCTGAAACACAAACAGCAGCATCTCTCCATGTTTCCATT 1089  
|||||

QY 121 GGCTCAGGATCATGCTGATTTAGTTACTCTGACCCAGCATCATGCTGATACGCGCAAG 180  
|||||  
Db 1090 GGCTCAGGATCATGCTGATTTAGTTACTCTGACCCAGCATCATGCTGATACGCGCAAG 1149  
|||||

QY 181 AACAAATACCGCTGATGAGTTACGCTGGAGAAAATAGAGCCCTTCCAAATGTGAGCACA 240  
|||||  
Db 1150 AACAAATACCGCTGATGAGTTACGCTGGAGAAAATAGAGCCCTTCCAAATGTGAGCACA 1209  
|||||

QY 241 GTATCCACATCTCTGACCTCCCGGAGAACGAAAGGAGAGCGTGCGAGTTCCAGTGG 300  
|||||  
Db 1210 GTATCCACATCTCTGACCTCCCGGAGAACGAAAGGAGAGCGTGCGAGTTCCAGTGG 1269  
|||||

QY 301 AAACAGGACAGCTGCGAGTGGGTGAGGTGATGAGCGCTGCTGGGCCCTG 351  
|||||  
Db 1270 AAACAGGACAGCTGCGAGTGGGTGAGGTGATGAGCGCTGCTGGGCCCTG 1320  
|||||

RESULT 3  
AB049473  
LOCUS AB049473 11187 bp mRNA linear ROD 01-DEC-2001  
DEFINITION Rattus norvegicus mRNA for reelin, complete cds.  
ACCESSION AB049473  
VERSION AB049473.1 GI:17221617  
KEYWORDS  
SOURCE Rattus norvegicus juvenile cerebellum cDNA to mRNA.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae;  
Rattus.

REFERENCE 1  
Kikkawa, S. and Terashima, T.  
Rat reelin (Reln) complete CDS  
Published Only in Database (2001)  
2 (bases 1 to 11187)  
Kikkawa, S. and Terashima, T.  
Direct Submission  
Submitted (29-SEP-2000) Satoshi Kikkawa, Kobe University School of  
Medicine, Department of Anatomy; Chuo-ku, Kusunoki-cho 7-5-1, Kobe,  
Hyogo 650-0017, Japan (E-mail: skikkawa@med.kobe-u.ac.jp,  
Tel:81-78-382-5325, Fax:81-78-382-5328)  
Location/Qualifiers  
1. .11187  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/tissue\_type="cerebellum"  
/dev\_stage="juvenile"  
1. .11187  
/gene="Reln"  
341..10729  
/gene="Reln"  
/codon\_start=1  
/product="reelin"  
/protein\_id="BAB78470.1"  
/db\_xref="GI:17221618"  
/translation="MERCWAPRTLVALLVLLLLTLRARATGYPRSPFFFLCETH  
HGELEGDEQGEVLISLHAGNPTVYVQGEYHVHTISTFDGLLVLTGLYSTSIQS  
SQSGGSAFGIMSDHQFNQFVSVASHVSHLPTNLNLFVMIAPAGGCVNFM  
ATATHRGQVFKDALAQOLCEOGAPTEAYSHLAEIHSDSVILRDDPDSYHOLELN  
NTWACSNCDTGEQGTLMHGNATFCPEYGPRLTNTYLTNTTASLOFSIGSGSCR  
FSYSDPSIIVSYAKNNTADWIOLEKIRAPSNVSTIIHILYLPEDAKGENVQFQWKQDS

LHVGEVYEAACWALDNILVNSAHROVILLEDSDLPDVTGNWLFEPGATVHKHSCQSDGNA  
IYFHNESQLNFAATRDVDLSTEDIQEQWSEEPESQPTGNDILGNVGVSEGTIESG  
LSLVFLKDEKLCPTMDTGTGNLRFYFAMGCTCDGDSHENDVLLAKLEGKKEH  
IADLTLSYSSYKVPVLVSVINPELQTPATKFLRQKNQHNQNNVAFVPHLPIL  
PMTSMHIOFSINLGCSTHQPNSVLESTNHRGWSLLHTECLPEICAGPHLPHST  
IYSSNYSGNRVPTLPNAALTROTIRWOTGPTILGNMWAIDNVIYIPGSLKFCFSG  
RGCTRHGCKCDPFGSPACEMASOTFPMFISESPGSSRLSSYHNFIYSIRAEVSGC  
GVLAGSKALVFNKDGRRQLITSFLOSSSRFLQFTLRGSKSVLSTCRADPDQGGVYL  
LHYSDNGITWKLLEHYSYLNHYHEPRIISVELPDDAKQFQIQRWMPQVHSSOGEDVM  
AIDEILMTSVLNSISLSTLEVLVETSLGFLYIGNIQPYCGHDWTLCTFGDSKLASSM  
RYVEQSMOIGASYMIQFSLVMGCGKYTPHMDNOVKLEYSTNHGILYTHADWDQCLPS  
MPCOEFTSASIVYHASEFTQWRVTVILPKOTSGATFRWQSOYTHADWDALDDIY  
IGQCPNMGCSHGSCDHGVCRCDOGYGTECHPEALPSTIMSDPENPSSWSDMQEY  
IGGEVVKPQGGCVVSSGSLYFSRAGKRLQVSWDLDTSWDFVQFYIQIGESASQCP  
KPDREGEVLLQYNNNGQVQWHLAEYFDEKPRFVLELPAAKATCTCFRFRMWP  
VSGEDYDQWAVDDIIILSEKQKQVPPVNPPTLPQNFYEKPAFDIPINQMSVWMLAN  
EGMAKNDSCATTPSAMVEGKSDGRFATRLTLTKPGYVLAQFKLNGICGASFSJAT  
VLLOYSHDAGMSWFLVKEGCFPASAGCKEGNSRELSEPTVYTGDFEWMRTVTAIP  
RSIASSKTRFRTOESSOKNVPFGLDGVIYISEPCPSYCSHGDCISGVCCDLGYT  
AAQGTCSNIPNHSEMFDFEGLSPWTKISGGQVGTGCGTSLSDGRSLYFNLGKRE  
ARTVPLDTRNIRLVQFYIQIGSKTSGITCIKPRARNEGLVQVYSDNGILMHLRELD  
FLSFLPQIISIDLPREKATPATFRWQPHQKHAQWALDDVLIGVNDSSQTFQD  
KFGSDIDLOANNYRIQGGQVDDICLSMDTALIFTENIGKPRYAEWDFHVSASFLOF  
DMSGCKSPSATHSVQLOYSLNKGMDHPVTEECVPPTIGCVHYTESSTYTSERON  
WRVTVYLPATNSPTRFRWTOANYTGADAWAIDNVLASGCPWLCSGRIGCDSGR  
CYCDRGFGPFCVPVPLPSILKDPFNGNLHFDLPEVYGAERNGNLNGETIRKGTSLI  
FKEGELRMLISRDLDCTNMYQVSLRFTAKGTPEHSHILOSLSNGTNGVTSWGLDEF  
YFPQTTSLFINVPLPYSQATNATFRWYNNKGKEIIDDIIDGDLNNPVM  
LLDTDFGPREDNWFFYPGGNIGLYCPSYSGKAPEDSASVNFVSIHGSITTRDLS  
VNPNTIIOFEINVGSTDSADPVRLEFSRDFGATWHLPLCYHSSSLVSLCSTE  
HNPSTIYAGTTQGHRRREVHFGKLHLCGSVRFRWYQGFYPAGSOPVTWALDNVYIGP  
QCEEMCCGHCVCNGTKICDPYSGPTCKISTKNPDFLKDDEFGEGSDRELLMSSGG  
KPSRKGLISSGNLFFNEDGLMLVTRDLDSHARFQVFRMLGCKGKGPVPRSPFY  
LQYSLNGLSWSLLQEFVLSNSNVYRIALEMPLKARSGSTRLWMPQSPNGRPOV  
PWVDOILIGNISGTVLEDDFSLDSRWLLHPGCTKMPVCGTGDALVTEKATST  
RYVVTDDIAVNEDSFLOIDFAASCVTDSYAEILEYSDVLGSLHPLVRDCLPTNVE  
CSRYHLQRLTVSDTFNKWTRITLPLPAYTSOATFRWHPQKQWMAIDNVIYIG  
DCLDMCSGHGRCIOGSCVDEQWGLGYCDEPETSPLTQKDNFRAPSNQWLVYIG  
GKLSVCGAVASGLALHFGSGCSRLVLDNLNNAEIOFYFVGLCLTTPNNRQVY  
LLEYSVNGISGTVLMEIFYDQYKSPGVNILLPPDAKEIGTRFRWMPQHDQDND  
WALDNVLSITWADQRTVMDLTSSAPVQHERSPADAPGVRIAFDMFEDKTSVNEH  
WYHDDCTVERFCDSPGVMLCGSHDREVYAVTHDTPENNIMQFKISVCKVPEK  
IAQNOIHVQFSDFGVSWSLVPOCLPADPKCSGTVSOPSVFEPFKGWIRIYPLPES  
LMGNVRFYQKISDVQWAINFYLGPCLDNCGHGDCLKEQCIDCPYSGPHCYL  
THLTFLKEREDSEIEKPDLMWLEGGSTECGGLAENTALTALYFGSTVRAQITQDL  
DLRGAKFLWGRIGSENNMTSCHRPKRGVLLDSDGGITWTLLHEMDFQYIS  
VRHDYILLPEGALTNTLRWQPPFVISNGLVSVGERAQWALDNILIGAEINPQSL  
VDTFDDGSSHEENNSFYNAVYTAGFCGNPSFHLVWPNKKDKDTHNALSSRELIIQ  
GYMPOFKIYVGCATSCDGLHSVMEYTKDARSQVLTQVOTCLPSSNSIGCSPPQF  
HEATIIYAVNSSSWRKITQLPDHVSSTATQFRWIKGEEERKQSWAIDHVTIGERCP  
RLCSGHGYCTGAVCIDESFGQDDCSVFESHELPSYIKDNFESARVTEANWETIOGGA  
IGSGGQLAPYAGHDSLYFNGCQIRQAATKPLDLTRASKIMFVLQISTGTAQDSCNSD  
LSGPHTVDKAVLLQYSVNGNITWVIAQHPKDFTAQARVSYNVLPEARMKGVLLRW  
OPRHNGTGHQWALDHVEVLVSTRKNYMMNFSRQHLRHFYNNRRRSLRRYP"

BASE COUNT 2645 a 3029 c 2930 g 2583 t  
ORIGIN

Query Match 85.4%; Score 299.8; DB 10; Length 11187;  
Best Local Similarity 90.9%; Pred. No. 2.5e-80;  
Matches 319; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1 GAGCAGTGTGGCACCACCATCATGCGAATGCTGTGACCTTCTGTGAGCGCGTACGGCCCT 60  
|||  
Db 1031 GAACAGTGTGGCACCACCATCATGCGAATGCTGTGACCTTCTGTGAGCGCATATGCTCT 1090  
|||

QY 61 CGAGAGCTGACACACACATGCTGAAACACAAACAGCAGCATCTCTCCATGTTTCCATT 120  
|||

Db 1091 CGAGAGTGTGACTACCAACATACCTGAACACGACAAACAGCATCCCGTCTCCAGTTCTCCATT 1150  
|||

QY 121 GGCTCAGGATCATGCTGATTTAGTTACTCTGACCCAGCATCATGCTGATACGCGCAAG 180  
|||

Db 1151 GGCTCAGGATCATGCTGATTTAGTTACTCTGACCCAGCATCATGCTGATACGCGCAAG 1210  
|||



VVSTDAVNEDSFLQIDFAASCVTDSCYAIELEYSDVLGLSWHPLVRDCLPTNVECS  
RYHLQRLVTFNKNWTRITLPLPYTSQATFRFRWHPAPDFKQOTQWALDINVIYGDG  
CDMCSGHRGCTQGCNCDQDEOGLGYCDPSETSLPTQLKDNFNRPSSQNLITVANGGK  
LSTVCGASGALHESGCSRLLVTLNLTNPDAEFOFYFMYGCLITPNNRQGVLL  
EYVNGGKITWNLLEIFVDOYKPGFVNILLPPDAKEIATFRWMPRHDJDDNDWA  
IDNVLISGADORTVMDLTFSSAPVPOHSPADAGPVGRITAFDMEDKTSVNEHWL  
FHDGCTVRECDSPDGVMLCGSHDREYVAVTHDLTPTEGIMQKISVGVKYSKIA  
QNIHVQISTDRGVSNNLYPOCLPADPKCSGVSQSPSVFFPTGKWKRTIYPLPESLV  
GNPVRFRYKYSMDMOWAIDNLYLPGCLDNRGHGDCLEQICDIPGSGPNCYLTH  
TLKTLFKERFDEEEKPDLMWSLEGSTCEGILAEIDTALYFGGTVRQAVTDOLDL  
RGAFTLQWGRIGSNNTSCHRPICRKEGVLLDSTGGITWTLLEHMDYOKYISVR  
HDYILLPDAINTNTRLMWOPFVINSIGVSSVERAOWALDNLIGGAENPSOLVD  
TDDDESHENWSPNAPVATAGCGNPSPHLYPNKKKDKKTHNALSRELLIQPGY  
MMQFKIVYCEATCGDLHSVMEYTKDARSDSWOLVQTCPLSSNSIGCSPPFQHE  
ATINYSNVSKRRTIQLPDHVSSATQFRWIQKEETEKOQSWAIDHVIYGEACPKL  
CSGHGYCTGATCIDESFGQDCSVFSDLPSTIKDNFESARVETANNETIQGVIG  
SCGGLAPYAGHDSLYFNGCQIRQAATPLDLTRASKIMFVLIQIGMSQTDSCSDLS  
GPHAVDKVLLVSYNNNGITMHWIAHQPKDFTQAOVSYNVPLEARMKGVLLRWQOP  
RHNGTGHQWALDHEVVLVSTRKONYMMNFSROHGLRHFYNNRRSLRRYP"

BASE COUNT 3014 a 2696 c 2753 g 3116 t 1 others  
ORIGIN

Query Match 75.1%; Score 263.6; DB 9; Length 11580;  
Best Local Similarity 84.6%; Pred. No. 3e-69;  
Matches 296; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 1 GAGCAGTGTGGCACCATCATGCATGCAATGCTGTCACTTCTGTGAGCGGTACGGCCCT 60  
DB 860 GAACAGTGTGGCGGATTTATGATGCAATGCGTCACTTCTGTGAACCATATGCGCCA 919  
QY 61 CGAGAGTGTACCACCATGCTCTGAACACAAACAGCATCTGTCTCCAGTTTCCATT 120  
DB 920 CGAGAACTGATTTACACAGCCCTTATACAAACAGCTTCTGTCTCCAAATTTCCATT 979  
QY 121 GGCTCAGGATCATGTGCAATTTAGTTACTCTGACCCAGCATCTGTGTCATACGCGCAAG 180  
DB 980 GGCTCAGGTTTCATGTCGCTTTAGTTATTTACAGACCCAGCATCATCGTGTATATGCGCAAG 1039  
QY 181 ACAATACCGCTGATTTGATTCAGCTGGAGAAAATTTAGAGCCCTTCCAAATGTGAGCACA 240  
DB 1040 AATAACTCTGCGGACCTGGATTCAGCTAGAGAAAATTTAGAGCCCTTCCAAATGTGAGCACA 1099  
QY 241 GTCATCCACATCTGTACCTCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGTGG 300  
DB 1100 ATCATCCATATCTCTACCTTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGTGG 1159  
QY 301 AAGACAGACGCTGCGAGTGGGTGAGGTGTATGAGGCGCTGCTGGGCCCT 350  
DB 1160 AAGCAGGAAAATCTTCTGTAGTGAAGTGTATGAAGCCTGCTGGGCCCT 1209

RESULT 6  
G30936  
LOCUS SW552926 Eric D. Green Homo sapiens STS genomic, sequence tagged  
DEFINITION site, 11580 bp DNA linear STS 28-SEP-1998

ACCESSION G30936  
VERSION G30936.1 GI:1923209  
KEYWORDS STS.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 11580)  
Bouffard,G.G., Iyer,L.M., Idol,J.R., Braden,V.V., Cunningham,A.F.,  
Weintraub,L.A., Mohr-Tidwell,R.M., Peluso,D.C., Fulton,R.S.,  
Leckie,M.P. and Green,E.D.  
A collection of 1814 human chromosome 7-specific STSs  
Genome Res. 7 (1), 59-64 (1997)

JOURNAL  
MEDLINE 97189344  
PUBMED 9037602  
REFERENCE 2 (bases 1 to 11580)  
AUTHORS Green,E.D.

TITLE  
JOURNAL  
COMMENT

Human chromosome 7 STSs (1997)  
Unpublished (1997)  
On Apr 3, 1997 this sequence version replaced gi:1706935.  
Synonyms: RELN  
GDB\_DSEG: RELN  
Contact: Eric D. Green  
Genome Technology Branch  
National Human Genome Research Institute/NIH  
49 Convent Dr., MSC4431, Bldg. 49, Rm. 2A08, Bethesda, MD 20892  
Tel: 3014020201  
Fax: 3014024735  
Email: egreen@nih.nih.gov  
Primer A: CTCACCTTCTGAGGAGCGCAA  
Primer B: AAGCAGTGTGCTGTGTCACCTG  
STS size: 183  
PCR Profile:  
Presoak: 0 degrees C for 0.00 minute(s)  
Denaturation: 92 degrees C for 0.17 minute(s)  
Annealing: 55 degrees C for 1.00 minute(s)  
Polymerization: 72 degrees C for 1.00 minute(s)  
PCR Cycles: 35  
Thermal Cycler: PerkinElmer 9600  
Protocol:  
Template: 30-100 ng  
Primer: each 1 uM  
dNTPs: each 200 uM  
Taq Polymerase: 0.05 units/ul  
Total Vol: 10 ul

Buffer: MgCl2: 1.5 mM  
KCl: 100 mM  
Tris-HCl: 10 mM  
NH4Cl: 5 mM  
pH: 8.6

This STS was developed from sequence determined by another investigator. See GenBank record: U79716 For additional information about the NHGRI chromosome 7 mapping project, see <http://www.nhgri.nih.gov/DIR/FTP/CH7>. Also see Genomics 11:548-64 (1991) [MUID=92128937].

FEATURES  
Location/Qualifiers  
1..11580  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="7"  
1..11580  
/clone\_lib="Eric D. Green"  
1112..1294  
/gene="RELN"  
1112..1134  
/gene="RELN"  
/gene="RELN"  
complement(1272..1294)  
primer\_bind 3014 a 2696 c 2753 g 3116 t 1 others  
BASE COUNT 3014 a 2696 c 2753 g 3116 t 1 others  
ORIGIN

Query Match 75.1%; Score 263.6; DB 11; Length 11580;  
Best Local Similarity 84.6%; Pred. No. 3e-69;  
Matches 296; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 1 GAGCAGTGTGGCACCATCATGCATGCAATGCTGTCACTTCTGTGAGCGGTACGGCCCT 60  
DB 860 GAACAGTGTGGCGGATTTATGATGCAATGCGTCACTTCTGTGAACCATATGCGCCA 919  
QY 61 CGAGAGTGTACCACCATGCTCTGAACACAAACAGCATCTGTCTCCAGTTTCCATT 120  
DB 920 CGAGAACTGATTTACACAGCCCTTATACAAACAGCTTCTGTCTCCAAATTTCCATT 979  
QY 121 GGCTCAGGATCATGTGCAATTTAGTTACTCTGACCCAGCATCTGTGTCATACGCGCAAG 180  
DB 980 GGCTCAGGTTTCATGTCGCTTTAGTTATTTACAGACCCAGCATCATCGTGTATATGCGCAAG 1039



Db	2	ACCAATGGCGCTTAACACAACTACTGCATCCGTCCTTCAGTTTTCCCTTGGGATCAGGTTCCG	61		
QY	133	TGTCGATTTAGTTACTCTGACCCAGCATCACTGTGTGCATACGCCAAGAACAAATACCGCT	192		
Db	62	TGCCGCTTTAGTTACTCAGATCCAGCATCAGGTCTCTACAGCAAGAACAGCTCTGCA	121		
QY	193	GATTGATTCAGCTGGAGAAAATAGAGCCCTTCCAAATGTGACGACAGTATCCACATC	252		
Db	122	GATTGGACTCAGCTGGAGAAAATAGTCTCTCTCCAAATGTGACGACGATCATACATC	181		
QY	253	CTGTACTCTCCCGAGGAAGCCAAAGGGAGAGCGTGCAGTTCCAGTGGAAACAGGACAGC	312		
Db	182	TTGTACTCTCTGAGGATGCTAAAGGGAGAGATGTCATTTTCTAGTGGAAACAGGACTAT	241		
QY	313	CTCGAGTGGGTGAGGTGTATGAGCCCTGCTGGGCCCTG	351		
Db	242	CTTCATGCTGGAGAGTGTATGAAGCCTGTTGGGCATTG	280		
RESULT 9					
AC121878/c					
LOCUS	AC121878	163985 bp	DNA linear		
DEFINITION	Mus musculus clone RP24-131C8, complete sequence.				
ACCESSION	AC121878				
VERSION	AC121878.1	GI:21039993			
KEYWORDS	HTG.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
REFERENCE	1 (bases 1 to 163985)				
AUTHORS	McPherson,J.D. and Waterston,R.H.				
TITLE	The sequence of Mus musculus clone				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 163985)				
AUTHORS	McPherson,J.D. and Waterston,R.H.				
TITLE	Direct Submission				
JOURNAL	Submitted (21-MAY-2002) Genome Sequencing Center, 4444 Forest Park				
REFERENCE	3 (bases 1 to 163985)				
AUTHORS	McPherson,J.D. and Waterston,R.H.				
TITLE	Direct Submission				
JOURNAL	Submitted (21-JUN-2002) Genome Sequencing Center, 4444 Forest Park				
COMMENT	Parway, St. Louis, MO 63108, USA				
----- Genome Center -----					
Center: Washington University Genome Sequencing Center					
Center code: WUGSC					
Web site: http://genome.wustl.edu/gsc/index.shtml					
Contact: submissions@watson.wustl.edu					
----- Project Information -----					
Center project name: M.BB0131C08					
-----					
FEATURES					
source					
1..163985					
/organism="Mus musculus"					
/db_xref="taxon:10090"					
/clone="RP24-131C8"					
BASE COUNT	50392	a	33960 c 32422 g 47211 t		
ORIGIN					
Query Match	38.5%	Score 135;	DB 10; Length 163985;		
Best Local Similarity	100.0%;	Pred. No. 1.1e-29;			
Matches 135;	Conservative	0; Mismatches	0; Indels 0; Gaps		
QY	217	AGAGCCCTTCCAAATGTGACGACAGTATCCACATCTCTGTACCTCCCGAGGAAGCCAAA	276		
Db	15382	AGAGCCCTTCCAAATGTGACGACAGTATCCACATCTCTGTACCTCCCGAGGAAGCCAAA	15323		
QY	277	GGGAGAGCGTGCAGTTCAGTGGAAACAGGACAGCTGCGAGTGGGTGAGGTATGAG	336		
Db	15322	GGGAGAGCGTGCAGTTCAGTGGAAACAGGACAGCTGCGAGTGGGTGAGGTATGAG	15263		





```

Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokkwo, S., Oguh, M., Okwuonu, G.,
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,
Scherer, S., Scott, G., Shen, H., Shooshtari, N., Slisson, I.,
Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 126130)
Worley, K.C.
Direct Submission
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Bay Plaza, Houston, TX 77030, USA
3 (bases 1 to 126130)
Worley, K.C.
Direct Submission
Submitted (11-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Bay Plaza, Houston, TX 77030, USA
On Jul 10, 2002 this sequence version replaced gi:20975941.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GDPS
Center clone name: CH230-10G13
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 70934 bases at least Q40
Consensus quality: 75985 bases at least Q30
Consensus quality: 80731 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 59 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1382: contig of 1382 bp in length
* 1383 1482: gap of unknown length
* 1483 3046: contig of 1564 bp in length
* 3047 3146: gap of unknown length
* 3147 4489: contig of 1343 bp in length
* 4490 4589: gap of unknown length
* 4590 5786: contig of 1197 bp in length
* 5787 5886: gap of unknown length
* 5887 7115: contig of 1229 bp in length
* 7116 7215: gap of unknown length
* 7216 8771: contig of 1556 bp in length
* 8772 8871: gap of unknown length
* 8872 10260: contig of 1389 bp in length
* 10261 10360: gap of unknown length
* 10361 11370: contig of 1010 bp in length
* 11371 11470: gap of unknown length
* 11471 12710: contig of 1240 bp in length
* 12711 12810: gap of unknown length
*
* 12811 14180: contig of 1370 bp in length
* 14181 14280: gap of unknown length
* 14281 16048: contig of 1768 bp in length
* 16049 16148: gap of unknown length
* 16149 17569: contig of 1421 bp in length
* 17570 17669: gap of unknown length
* 17670 18920: contig of 1151 bp in length
* 18921 18920: gap of unknown length
* 18921 20209: contig of 1289 bp in length
* 20210 20309: gap of unknown length
* 20310 21633: contig of 1324 bp in length
* 21634 21733: gap of unknown length
* 21734 23269: contig of 1536 bp in length
* 23270 23369: gap of unknown length
* 23370 24647: contig of 1278 bp in length
* 24648 24747: gap of unknown length
* 24748 25800: contig of 1053 bp in length
* 25801 25900: gap of unknown length
* 25901 27017: contig of 1117 bp in length
* 27018 27117: gap of unknown length
* 27118 29183: contig of 2086 bp in length
* 29184 29283: gap of unknown length
* 29284 30315: contig of 1032 bp in length
* 30316 30415: gap of unknown length
* 30416 32086: contig of 1671 bp in length
* 32087 32186: gap of unknown length
* 32187 33395: contig of 1209 bp in length
* 33396 33495: gap of unknown length
* 33496 35663: contig of 2168 bp in length
* 35664 35763: gap of unknown length
* 35764 38044: contig of 2281 bp in length
* 38045 38144: gap of unknown length
* 38145 39845: contig of 1701 bp in length
* 39846 39945: gap of unknown length
* 39946 41982: contig of 2037 bp in length
* 41983 42082: gap of unknown length
* 42083 43743: contig of 1661 bp in length
* 43744 43843: gap of unknown length
* 43844 45904: contig of 2061 bp in length
* 45905 46004: gap of unknown length
* 46005 47667: contig of 1663 bp in length
* 47668 47767: gap of unknown length
* 47768 49833: contig of 2066 bp in length
* 49834 49933: gap of unknown length
* 49934 51434: contig of 1501 bp in length
* 51435 51534: gap of unknown length
* 51535 53778: contig of 2244 bp in length
* 53779 53878: gap of unknown length
* 53879 55552: contig of 1674 bp in length
* 55553 55652: gap of unknown length
* 55653 57670: contig of 2018 bp in length
* 57671 57770: gap of unknown length
* 57771 59630: contig of 1860 bp in length
* 59631 59730: gap of unknown length
* 59731 61147: contig of 1417 bp in length
* 61148 61247: gap of unknown length
* 61248 63995: contig of 2748 bp in length
* 63996 64095: gap of unknown length
* 64096 66294: contig of 2199 bp in length
* 66295 66394: gap of unknown length
* 66395 68794: contig of 2400 bp in length
* 68795 68894: gap of unknown length
* 68895 71533: contig of 2639 bp in length
* 71534 71633: gap of unknown length
* 71634 73469: contig of 1836 bp in length
* 73470 73569: gap of unknown length
* 73570 77054: contig of 3485 bp in length
* 77055 77154: gap of unknown length
* 77155 79005: contig of 1851 bp in length
* 79006 79105: gap of unknown length
* 79106 81294: contig of 2189 bp in length
* 81295 81394: gap of unknown length
* 81395 83864: contig of 2470 bp in length

```

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

```
* 83865 83964: gap of unknown length
* 83965 86039: contig of 2075 bp in length
* 86040 86139: gap of unknown length
* 86140 88646: contig of 2507 bp in length
* 88647 88746: gap of unknown length
* 88747 91823: contig of 3077 bp in length
* 91824 91923: gap of unknown length
* 91924 94369: contig of 2446 bp in length
* 94370 94459: gap of unknown length
* 94470 96886: contig of 2417 bp in length
* 96887 96986: gap of unknown length
* 96987 100679: contig of 3693 bp in length
* 100680 100779: gap of unknown length
* 100780 104171: contig of 3392 bp in length
* 104172 104271: gap of unknown length
* 104272 106760: contig of 2489 bp in length

Query Match 33.0%; Score 115.8; DB 2; Length 126130;
Best Local Similarity 91.1%; Pred. No. 7.5e-24;
Matches 123; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 217 AGAGCCCTTCCAAATGTGAGCAGTCATCCACATCTGTACCTGCCGAGGAGGACCAA 276
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 92827 AGAGCCCTTCCAAATGTGAGCAGTCATCCACATCTGTACCTGCCGAGGAGGACCAA 92886

QY 277 GGGGAGAGCGTCGAGTTCCTAGTGGAAACAGCAGCCTGCGAGTGGGTGAGGTGTATGAG 336
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 92887 GGGGAGAGCGTCGAGTTCCTAGTGGAAACAGCAGCCTGCGAGTGGGTGAGGTGTATGAG 336

QY 337 GCCTGCTGGGCCCTG 351
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 92947 GCCTGCTGGGCCCTG 92961

RESULT 12
AC128022/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-525L20, *** SEQUENCING IN PROGRESS
***, 97 unordered pieces.
ACCESSION AC128022
VERSION AC128022.1 GI:21908605
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 202764)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,
Homsí,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlssoon,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
```

```
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Mosser,M., Neal,D., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu.L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojuboan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 202764)
Worley,K.C.
Direct Submission
Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor-College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: KAIV
Center clone name: CH230-525L20
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 113069 bases at least Q40
Consensus quality: 119970 bases at least Q30
Consensus quality: 125656 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 97 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1244: contig of 1244 bp in length
* 1245 1344: gap of unknown length
* 1345 2346: contig of 1002 bp in length
* 2347 2446: gap of unknown length
* 2447 3813: contig of 1367 bp in length
* 3814 3913: gap of unknown length
* 3914 5101: contig of 1188 bp in length
* 5102 5201: gap of unknown length
* 5202 6724: contig of 1523 bp in length
* 6725 6824: gap of unknown length
* 6825 8048: contig of 1224 bp in length
* 8049 8148: gap of unknown length
* 8149 9212: contig of 1064 bp in length
* 9213 9312: gap of unknown length
* 9313 10432: contig of 1120 bp in length
* 10433 10532: gap of unknown length
* 10533 11890: contig of 1358 bp in length
* 11891 11990: gap of unknown length
* 11991 13068: contig of 1078 bp in length
* 13069 13168: gap of unknown length
* 13169 14486: contig of 1318 bp in length
* 14487 14586: gap of unknown length
* 14587 15996: contig of 1410 bp in length
* 15997 16096: gap of unknown length
```

\* 16097 17534: contig of 1438 bp in length  
\* 17335 17634: gap of unknown length  
\* 17635 18703: contig of 1069 bp in length  
\* 18704 18803: gap of unknown length  
\* 18804 19884: contig of 1081 bp in length  
\* 19885 19984: gap of unknown length  
\* 19985 21539: contig of 1555 bp in length  
\* 21540 21639: gap of unknown length  
\* 21640 22780: contig of 1141 bp in length  
\* 22781 22880: gap of unknown length  
\* 22881 23926: contig of 1046 bp in length  
\* 23927 24026: gap of unknown length  
\* 24027 25544: contig of 1518 bp in length  
\* 25545 25644: gap of unknown length  
\* 25645 26899: contig of 1255 bp in length  
\* 26900 26999: gap of unknown length  
\* 27000 28237: contig of 1238 bp in length  
\* 28238 28337: gap of unknown length  
\* 28338 29758: contig of 1421 bp in length  
\* 29759 29858: gap of unknown length  
\* 29859 31038: contig of 1180 bp in length  
\* 31039 31138: gap of unknown length  
\* 31139 32179: contig of 1041 bp in length  
\* 32180 32279: gap of unknown length  
\* 32280 33797: contig of 1517 bp in length  
\* 33797 33896: gap of unknown length  
\* 33897 35380: contig of 1484 bp in length  
\* 35381 35480: gap of unknown length  
\* 35481 37032: contig of 1552 bp in length  
\* 37033 37132: gap of unknown length  
\* 37133 38414: contig of 1282 bp in length  
\* 38415 38514: gap of unknown length  
\* 38515 39948: contig of 1434 bp in length  
\* 39949 40048: gap of unknown length  
\* 40049 41436: contig of 1388 bp in length  
\* 41437 41536: gap of unknown length  
\* 41537 43084: contig of 1548 bp in length  
\* 43085 43184: gap of unknown length  
\* 43185 44940: contig of 1756 bp in length  
\* 44941 45040: gap of unknown length  
\* 45041 46186: contig of 1146 bp in length  
\* 46187 46286: gap of unknown length  
\* 46287 47989: contig of 1703 bp in length  
\* 47990 48089: gap of unknown length  
\* 48090 49842: contig of 1753 bp in length  
\* 49843 49942: gap of unknown length  
\* 49944 51128: contig of 1186 bp in length  
\* 51129 51228: gap of unknown length  
\* 51229 52823: contig of 1595 bp in length  
\* 52824 52923: gap of unknown length  
\* 52924 54431: contig of 1508 bp in length  
\* 54332 54531: gap of unknown length  
\* 54532 56042: contig of 1511 bp in length  
\* 56043 56142: gap of unknown length  
\* 56143 57287: contig of 1145 bp in length  
\* 57288 57387: gap of unknown length  
\* 57388 58893: contig of 1506 bp in length  
\* 58894 60024: contig of 1031 bp in length  
\* 60025 60124: gap of unknown length  
\* 60125 62858: contig of 2734 bp in length  
\* 62859 62958: gap of unknown length  
\* 62959 64365: contig of 1407 bp in length  
\* 64366 64465: gap of unknown length  
\* 64466 66434: contig of 1969 bp in length  
\* 66435 66534: gap of unknown length  
\* 66535 68631: contig of 2097 bp in length  
\* 68632 68731: gap of unknown length  
\* 68732 70839: contig of 2108 bp in length  
\* 70840 70939: gap of unknown length  
\* 70940 73109: contig of 2070 bp in length  
\* 73110 74957: contig of 1848 bp in length

\* 74958 75057: gap of unknown length  
\* 75058 76816: contig of 1759 bp in length  
\* 76817 76916: gap of unknown length  
\* 76917 78679: contig of 1763 bp in length  
\* 78680 78779: gap of unknown length  
\* 78780 80056: contig of 1277 bp in length  
\* 80057 80156: gap of unknown length  
\* 80157 81312: contig of 1156 bp in length  
\* 81313 82922: contig of 1510 bp in length  
\* 82923 83023: gap of unknown length  
\* 83024 84180: contig of 1158 bp in length  
\* 84181 84280: gap of unknown length  
\* 84281 85849: contig of 1569 bp in length  
\* 85850 85949: gap of unknown length  
\* 85950 87612: contig of 1663 bp in length  
\* 87613 87712: gap of unknown length

Query Match 33.0%; Score 115.8; DB 2; Length 202764;  
Best Local Similarity 91.1%; Pred. NO. 8.3e-24;  
Matches 123; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 217 AGAGCCCTTCCAAATGTGAGCAGATCCACATCCTGTACCTCCCGAGAGCCAAA 276  
|||||  
Db 187315 AGAGCCCTTCCAAATGTGAGCAGATCCACATCCTGTACCTCCCGAGAGCCAAA 187256  
|||||

QY 277 GGGGAGAGCGTGCAGTCCAGTGGAAACAGACAGCGCTGCGAGTGGGTGATGAG 336  
|||||  
Db 187255 GGGGAGAGCGTGCAGTCCAGTGGAAACAGACAGCGCTGCGTGGGTGATGAG 187196  
|||||

QY 337 GCCTGCTGGGCCCTG 351  
|||||  
Db 187195 GCCTGCTGGGCCCTG 187181  
|||||

RESULT 13  
AC124933/C  
LOCUS  
DEFINITION Rattus norvegicus clone CH230-466N17, \*\*\* SEQUENCING IN PROGRESS  
AC124933  
AC124933  
VERSION AC124933.2 GI:21952716  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 183641)  
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,  
Barbaria,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,  
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,  
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
Chen,G., Chen,K., Chen,Z., Chowdhry,I., Christopoulos,C.,  
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,  
Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,  
Homsif,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,  
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
Karlssoon,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,  
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,

```

Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokkwo,S., Ogih,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,
Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wlarczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 183641)
Worley,K.C.
Direct Submission
Submitted (20-JUN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 183641)
Worley,K.C.
Direct Submission
Submitted (31-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 24, 2002 this sequence version replaced gi:21490053.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: KAVI
Center clone name: CH230-466N17
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 111724 bases at least Q40
Consensus quality: 120970 bases at least Q30
Consensus quality: 128458 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 88 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1
* 1107: contig of 1107 bp in length
* 1108
* 1207: gap of unknown length
* 1208
* 2460: contig of 1253 bp in length
* 2461
* 2560: gap of unknown length
* 2561
* 4004: contig of 1444 bp in length
* 4005
* 4104: gap of unknown length
* 4105
* 5382: contig of 1278 bp in length
* 5383
* 5482: gap of unknown length
* 5483
* 6874: contig of 1392 bp in length
* 6875
* 6974: gap of unknown length
* 6975
* 8071: contig of 1097 bp in length
* 8072
* 8171: gap of unknown length
* 8172
* 9478: contig of 1307 bp in length
* 9479
* 9578: gap of unknown length
* 9579
* 11061: contig of 1483 bp in length
* 11062
* 11161: gap of unknown length

```

```

11162
12742: contig of 1581 bp in length
12842: gap of unknown length
12843
13843
14504: contig of 1662 bp in length
14604: gap of unknown length
14605
15734: contig of 1130 bp in length
15834: gap of unknown length
15835
16958: contig of 1124 bp in length
16959
17059
18004: contig of 1646 bp in length
18005
18805
20420: contig of 1616 bp in length
20421
20521
21648: contig of 1128 bp in length
21649
23177: contig of 1429 bp in length
23178
23277: gap of unknown length
23278
24760: contig of 1483 bp in length
24761
24860: gap of unknown length
24861
26759: contig of 1899 bp in length
26760
26859: gap of unknown length
26860
28399: contig of 1440 bp in length
28399
28399: gap of unknown length
28400
29592: contig of 1193 bp in length
29593
30870: contig of 1178 bp in length
30871
30970: gap of unknown length
30971
32642: contig of 1672 bp in length
32643
32742: gap of unknown length
32743
34386: contig of 1644 bp in length
34387
34486: gap of unknown length
34487
36064: contig of 1578 bp in length
36065
37398: contig of 1234 bp in length
37399
37499: gap of unknown length
37499
39608: contig of 2110 bp in length
39609
41480: contig of 1772 bp in length
41481
41580: gap of unknown length
41581
42759: contig of 1179 bp in length
42760
42859: gap of unknown length
42860
44711: contig of 1852 bp in length
44712
44811: gap of unknown length
44812
46208: contig of 1397 bp in length
46209
46308: gap of unknown length
46309
47767: contig of 1459 bp in length
47768
47867: gap of unknown length
47868
49227: contig of 1360 bp in length
49228
49327: gap of unknown length
49328
50797: contig of 1470 bp in length
50798
51928: contig of 1031 bp in length
51929
52028: gap of unknown length
52029
53105: contig of 1077 bp in length
53106
53205: gap of unknown length
53206
54546: contig of 1341 bp in length
54547
55820: contig of 1174 bp in length
55821
55920: gap of unknown length
55921
57635: contig of 1715 bp in length
57636
57735: gap of unknown length
57736
58942: contig of 1207 bp in length
58943
59042: gap of unknown length
59043
60341: contig of 1299 bp in length
60342
60441: gap of unknown length
60442
61693: contig of 1252 bp in length
61694
61793: gap of unknown length
61794
63315: contig of 1522 bp in length
63316
63416: gap of unknown length
63416
64571: contig of 1156 bp in length
64572
64671: gap of unknown length
64672
66083: contig of 1412 bp in length
66084
66183: gap of unknown length
66184
67516: contig of 1333 bp in length

```

```
* 67517 67616: gap of unknown length
* 67617 69820: contig of 2204 bp in length
* 69821 69920: gap of unknown length
* 69921 71261: contig of 1341 bp in length
* 71262 72775: gap of unknown length
* 71362 72775: contig of 1414 bp in length
* 72776 72875: gap of unknown length
* 72876 75298: contig of 2423 bp in length
* 75299 75398: gap of unknown length
* 75399 77094: contig of 1696 bp in length
* 77095 77194: gap of unknown length
* 77195 78834: contig of 1640 bp in length
* 78835 78934: gap of unknown length
* 78935 79989: contig of 1055 bp in length
* 79990 80089: gap of unknown length
* 80090 81774: contig of 1685 bp in length
* 81775 81874: gap of unknown length
* 81875 83255: contig of 1381 bp in length

Query Match      32.58; Score 114.2; DB 2; Length 183641;
Best Local Similarity 90.4%; Pred. No. 2.5e-23;
Matches 122; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 217 AGAGCCCTTCCATGTGAGCAGATCATCCATCTGTACCTCCCGAGGAGCCAAA 276
      ||||| ||||| || ||||| ||||| ||||| ||||| || ||||| |||||
Db 143722 AGAGCCCTTCCACGTAAGCACCATCATCCATCTCTACCTCTCTGAGGAGCCAAA 143663

Qy 277 GGGGAGCGTGCGAGTTCCAGTGGAAACAGGACAGCGCTGCGAGTGGGTGCTGTATGAG 336
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 143662 GGGGAGACGTCGAGTTCCTGAGTGGAAACAGGACAGCGCTGCTGTGGGTGCTGTACGAA 143603

Qy 337 GCCTGCTGGGCCCTG 351
      ||||| ||||| ||||| ||
Db 143602 GCCTGCTGGGCCCTG 143588

RESULT 14
HSAC000121/c
LOCUS      HSAC000121      93163 bp      DNA      linear      PRI 31-JAN-1997
DEFINITION Human BAC clone RG249A12 from 7q22, complete sequence.
ACCESSION AC000121
VERSION    AC000121.1 GI:1809222
KEYWORDS   HTG.
SOURCE     Homo sapiens.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 93163)
            Lamar, B., Wamsley, P. and Gibson, A.
            The sequence of H. sapiens BAC clone RG249A12
JOURNAL    Unpublished (1997)
REFERENCE  2 (bases 1 to 93163)
            Waterston, R.
            Direct Submission
            Submitted (31-JAN-1997)
JOURNAL    Genome Sequencing Center
            Department of Genetics, Washington University
            St. Louis, MO 63108, USA
            http://genome.wustl.edu/gsc
            e-mail: sapiens@watson.wustl.edu
```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

```

SOURCE INFORMATION:
This clone is from a release of the human BAC library. The library contains cloned DNA from a human sperm. For references see: Shizuya et al., Proc. Natl. Acad. Sci. 89:8794-8797 (1992); Kim et al., Genomics 34:213-218 (1996).
VECTOR: pBELO
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is H_RG352J05. Actual start of this clone is at base position 1 of H_RG249A12; actual end is at 93163 of H_RG249A12. This clone is part of an unanchored island, orientation is unknown.

This clone contains STS SWSS2926 (NID:gl706935).

FEATURES
      source
      1..93163
      Location/Qualifiers
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="7"
      /map="7q22"
      /clone="RG249A12"
      complement(1011..1309)
      /rpt_family="ALU"
      repeat_region
      2887..3178
      /rpt_family="ALU"
      misc_feature
      5965..6163
      /note="similar to C2H2-type zinc fingers PROSITE; PS00028"
      repeat_region
      complement(7118..7170)
      /rpt_family="L1"
      repeat_region
      9584..12339
      /rpt_family="L1"
      exon
      10207..11298
      /note="GRAIL prediction, score = 85"
      /evidence=not_experimental
      repeat_region
      complement(10583..11001)
      /rpt_family="L1"
      repeat_region
      12423..12714
      /rpt_family="ALU"
      repeat_region
      12715..12734
      /rpt_family="L1"
      repeat_region
      13365..13412
      /rpt_family="L1"
      repeat_region
      14286..14309
      /rpt_family="L1"
      repeat_region
      15942..15963
      /rpt_family="L1"
      gene
      complement(<17055..>84441)
      /gene="WUGSC:H_RG249A12.1"
      CDS
      complement(Join(<17055..17200,32792..33032,35849..35945,58079..58130,63050..63146,84365..>84441))
      /gene="WUGSC:H_RG249A12.1"
      /note="extracellular matrix protein; 97% similarity to mouse reelin U24703 (PID:g902487), partial CDS"
      /codon_start=-1
      /evidence=not_experimental
      /protein_id="AAB46357.1"
      /db_xref="GI:1809223"
      /translation="EIHSDSIILDRDFDSYHQLQNPNIWVECNCTGEOCGAIMHG
      NLEKFCPEYGPRLITTLGNTTASVLPQSIGSGCRFSYSDPSIILVYAKNNSADMI
      QVETKIRAPNSVTIIHILYLPEDAKGNVQFQWKQENLRVGEVYEACWALDNLIIINS
      AHRQVVLSDLPDVTGNLWLPFGATVKHSCQSDGNSIYPHGNESGFNFATTTROVDL
      STEDIQEWSEEFESQPTG"
      complement(17055..17252)
      exon
      /gene="WUGSC:H_RG249A12.1"
      /note="GRAIL prediction, score = 100"
      /evidence=not_experimental
      repeat_region
      21484..21521
      /rpt_family="L1"
      repeat_region
      complement(21962..21990)
      /rpt_family="L1"
      repeat_region
      24430..24466
```

```
repeat_region /rpt_family="L1"
24613..24654
repeat_region /rpt_family="L1"
25141..25180
repeat_region /rpt_family="L1"
26470..26761
misc_feature /rpt_family="ALU"
/genes="WUGSC:H_RG249A12.1"
complement(26762..26936)
repeat_region /note="match to Human cDNA, 5' end H56746 (NID:1005390)"
26937..26989
exon /rpt_family="L1"
27107..27238
/notes="GRAIL prediction, score = 82"
/coverage=not_experimental
repeat_region 27993..28465
repeat_region /rpt_family="L1"
28476..28496
repeat_region /rpt_family="L1"
complement(29663..29903)
exon /rpt_family="ALU"
complement(32793..33032)
/genes="WUGSC:H_RG249A12.1"
/coverage=not_experimental
repeat_region /note="GRAIL prediction, score = 90"
complement(33462..33752)
repeat_region /rpt_family="ALU"
37975..37994
repeat_region /rpt_family="L1"
40278..40311
repeat_region /rpt_family="L1"
complement(41921..42075)
repeat_region /rpt_family="ALU"
complement(42447..42734)
exon /rpt_family="ALU"
43126..43185
/notes="GRAIL prediction, score = 95"
/coverage=not_experimental
repeat_region complement(43411..43680)
repeat_region /rpt_family="ALU"
45704..45991
misc_feature /rpt_family="ALU"
46064..46434
/notes="match to Human cDNA H56097, 5' end (NID:g1012929)"
repeat_region complement(46604..46897)
misc_feature /rpt_family="ALU"
complement(47023..47412)
/genes="WUGSC:H_RG249A12.1"
/notes="match to Human cDNA, 3' end H59444 (NID:g1012276)"
58125..58163
repeat_region /rpt_family="L1"
58604..58758
repeat_region /rpt_family="L1"
59743..59765
repeat_region /rpt_family="L1"
60210..60516
exon /rpt_family="ALU"
complement(63051..63146)
/genes="WUGSC:H_RG249A12.1"
/notes="GRAIL prediction, score = 100"
/coverage=not_experimental
repeat_region 64242..64284
repeat_region /rpt_family="L1"
64445..64496
repeat_region /rpt_family="L1"
complement(67753..68011)
repeat_region /rpt_family="L1"
67986..68889
exon /rpt_family="L1"
complement(70725..70841)
/genes="WUGSC:H_RG249A12.1"
/notes="GRAIL prediction, score = 82"
/coverage=not_experimental
```

```
repeat_region 71911..71947
/rpt_family="L1"
complement(73631..73919)
repeat_region /rpt_family="ALU"
75073..75092
repeat_region /rpt_family="L1"
76743..82366
exon /rpt_family="L1"
78326..79402
/notes="GRAIL prediction, score = 83"
/coverage=not_experimental
exon 80586..81548
/notes="GRAIL prediction, score = 82"
/coverage=not_experimental
repeat_region complement(80962..81380)
repeat_region /rpt_family="L1"
complement(82581..82608)
misc_feature /rpt_family="L1"
83877..84447

Query Match 28.1%; Score 98.8; DB 9; Length 93163;
Best Local Similarity 83.6%; Pred. No. 1.1e-18;
Matches 112; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 217 AGAGCCCTTCCAAATGTGACACAGTCATCCACATCCTGTACCTCCCGAGGAGCCAAA 276
|||||
Db 33034 AGAGCCCTTCCAAATGTGACACAAATCATCCATATCCTCTACCTCTCTGAGGAGCCAAA 32975

QY 277 GGGGAGAGCGTCGAGTTCAGTGGAAACAGGACAGCGCTGGGAGTGGGTGATGAG 336
|||||
Db 32974 GGGGAGAATGTCCAAATTTCAAGTGGAAAGCAGGAAAATCTTCGTAGTGAAGTATGAA 32915

QY 337 GCCTGCTGGGCCCT 350
|||||
Db 32914 GCCTGCTGGGCCCT 32901

RESULT 15
AC041023/c
LOCUS AC041023
DEFINITION Homo sapiens clone RP11-356G3, WORKING DRAFT SEQUENCE, 20 unordered
pieces.
AC041023
VERSION AC041023.2 GI:8016758
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 185996)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-356G3
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 185996)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczyk,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihoval,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
```

Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A., and Zody, M.

TITLE  
JOURNAL

COMMENT

Submitted (11-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On May 22, 2000 this sequence version replaced gi:7534225.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L9601

Center clone name: 356\_G\_3

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 175155 bases at least Q40

Consensus quality: 180623 bases at least Q30

Consensus quality: 182712 bases at least Q20

Insert size: 188000; agarose-fp

Quality coverage: 4.5 in Q20 bases; agarose-fp

Quality coverage: 4.6 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 20 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 2941: contig of 2941 bp in length  
\* 2942 3041: gap of 100 bp  
\* 3042 5193: contig of 2152 bp in length  
\* 5194 5293: gap of 100 bp  
\* 5294 7050: contig of 1757 bp in length  
\* 7051 7150: gap of 100 bp  
\* 7151 10452: contig of 3302 bp in length  
\* 10453 10552: gap of 100 bp  
\* 10553 13859: contig of 3307 bp in length  
\* 13860 13959: gap of 100 bp  
\* 13960 17254: contig of 3295 bp in length  
\* 17255 17354: gap of 100 bp  
\* 17355 20523: contig of 3169 bp in length  
\* 20524 20623: gap of 100 bp  
\* 20624 25148: contig of 4525 bp in length  
\* 25149 25248: gap of 100 bp  
\* 25249 29676: contig of 4428 bp in length  
\* 29677 29776: gap of 100 bp  
\* 29777 34077: contig of 4301 bp in length  
\* 34078 34177: gap of 100 bp  
\* 34178 39009: contig of 4832 bp in length  
\* 39010 39109: gap of 100 bp  
\* 39110 45564: contig of 6455 bp in length  
\* 45565 45664: gap of 100 bp  
\* 45665 54035: contig of 8371 bp in length  
\* 54036 54135: gap of 100 bp  
\* 54136 63394: contig of 9259 bp in length  
\* 63395 63494: gap of 100 bp  
\* 63495 74946: contig of 11452 bp in length  
\* 74947 75046: gap of 100 bp  
\* 75047 88353: contig of 13307 bp in length  
\* 88354 88453: gap of 100 bp  
\* 88454 104853: contig of 16400 bp in length  
\* 104854 104953: gap of 100 bp  
\* 104954 123339: contig of 18386 bp in length  
\* 123340 123439: gap of 100 bp

\* 123440 154730: contig of 31291 bp in length  
\* 154731 154830: gap of 100 bp  
\* 154831 185996: contig of 31166 bp in length.

FEATURES

source

1. .185996

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="RP11-356G3"

/clone\_lib="RPCI-11 Human Male BAC"

1. .2941

misc\_feature

3042..5193

/note="assembly\_fragment"

misc\_feature

5294..7050

/note="assembly\_fragment"

misc\_feature

7151..10452

/note="assembly\_fragment"

misc\_feature

10553..13859

/note="assembly\_fragment"

misc\_feature

13960..17254

/note="assembly\_fragment"

misc\_feature

17355..20523

/note="assembly\_fragment"

misc\_feature

20624..25148

/note="assembly\_fragment"

misc\_feature

25249..29676

/note="assembly\_fragment"

misc\_feature

29777..34077

/note="assembly\_fragment"

misc\_feature

34178..39009

/note="assembly\_fragment"

misc\_feature

39110..45564

/note="assembly\_fragment"

misc\_feature

45665..54035

/note="assembly\_fragment"

misc\_feature

54136..63394

/note="assembly\_fragment"

misc\_feature

63495..74946

/note="assembly\_fragment"

misc\_feature

75047..88353

/note="assembly\_fragment"

misc\_feature

88454..104853

/note="assembly\_fragment"

misc\_feature

104954..123339

/note="assembly\_fragment"

misc\_feature

123440..154730

/note="assembly\_fragment"

misc\_feature

154831..185996

/note="assembly\_fragment"

misc\_feature

vector\_side:left"

vector\_side:right"

BASE COUNT 60145 a 34255 c 33646 g 56049 t 1901 others

ORIGIN

Query Match 15.3%; Score 53.8; DB 2; Length 185996;

Best Local Similarity 79.0%; Pred. No. 7.4e-05;

Matches 64; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 1 GACACGTGTGGCCACCATCATGCGCAATGCTGTGTCACCTTCTGTGAGCGGTACGGCCCT 60

|| ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||

Db 166372 GAACAGTGTGGCGCGATTATGTCATGCAATGCGCGTACCTTCTGTGAGCCATATGCCCA 166313

QY 61 CGAGAGCTGACACCATCATGCG 81

||||| ||| | |||

Db 166312 CGAGAACTGTAAGTATGTGC 166292

Search completed: November 6, 2002, 14:55:41

Job time : 3118 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

# OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 14:55:47 ; Search time 86 Seconds  
(without alignments)  
1359.737 Million cell updates/sec

Title: US-09-897-438B-1  
Perfect score: 351  
Sequence: 1 ggcagtggtgcaccatcat.....atgaggcctgctgggcccctg 351

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 310279 seqs, 166577418 residues

Total number of hits satisfying chosen parameters: 620558

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database : Published Applications, NA.\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	351	100.0	351	10	US-09-897-438B-1
2	263.6	75.1	11580	10	US-09-880-107-3436
C 3	98.8	28.1	499	10	US-09-864-761-14497
C 4	81.4	23.2	214	10	US-09-864-761-31037
C 5	53.8	15.3	413	10	US-09-864-761-11808
C 6	46.2	13.2	87	10	US-09-864-761-28379
C 7	31.6	9.0	168	10	US-09-878-574-14485
8	31.4	8.9	670	10	US-09-828-644-8
9	31.4	8.9	1536	12	US-10-050-726-1
10	31.4	8.9	1860	10	US-09-828-644-9
C 11	31.2	8.9	417	10	US-09-922-217-605
C 12	31.2	8.9	417	10	US-09-833-263-605
C 13	31.2	8.9	2532	12	US-10-124-598-4
14	30.6	8.7	2298	10	US-09-476-242-22
15	30.6	8.7	2298	10	US-09-476-242-23
16	30.6	8.7	2298	10	US-09-476-242-24
17	30.6	8.7	2310	10	US-09-476-242-3
18	30.6	8.7	2310	10	US-09-476-242-21
19	30.6	8.7	2316	10	US-09-476-242-4

20	30.6	8.7	2316	10	US-09-476-242-8	Sequence 8, Appli
21	30.6	8.7	2322	10	US-09-476-242-5	Sequence 5, Appli
22	30.6	8.7	2322	10	US-09-476-242-18	Sequence 18, Appl
23	30.6	8.7	2322	10	US-09-476-242-19	Sequence 19, Appl
24	30.6	8.7	2322	10	US-09-476-242-20	Sequence 20, Appl
25	30.6	8.7	2328	10	US-09-476-242-6	Sequence 6, Appli
26	30.6	8.7	2334	10	US-09-476-242-7	Sequence 7, Appli
27	30.6	8.7	2352	10	US-09-476-242-26	Sequence 26, Appli
28	30.6	8.7	2358	10	US-09-476-242-25	Sequence 25, Appl
29	30.6	8.7	2517	10	US-09-476-242-16	Sequence 16, Appl
30	30.6	8.7	2517	10	US-09-476-242-17	Sequence 17, Appl
31	30.6	8.7	2523	10	US-09-476-242-15	Sequence 15, Appl
32	30.6	8.7	2529	10	US-09-476-242-14	Sequence 14, Appl
33	30.6	8.7	2535	10	US-09-476-242-13	Sequence 13, Appl
34	30.6	8.7	2541	10	US-09-476-242-9	Sequence 9, Appli
35	30.6	8.7	2541	10	US-09-476-242-10	Sequence 10, Appl
36	30.6	8.7	2541	10	US-09-476-242-11	Sequence 11, Appl
37	30.6	8.7	2541	10	US-09-476-242-12	Sequence 12, Appl
C 38	30	8.5	485	10	US-09-864-761-10999	Sequence 10999, A
39	29.8	8.5	1017	10	US-09-777-430A-47	Sequence 47, Appl
40	29.6	8.4	2329	10	US-09-084-491A-1	Sequence 1, Appli
C 41	29.6	8.4	257	10	US-09-728-445-503	Sequence 503, App
42	29.4	8.4	677	10	US-09-738-973-448	Sequence 448, App
43	29.4	8.4	2657	10	US-09-880-107-2139	Sequence 2139, Ap
44	29.4	8.4	3522	10	US-09-815-242-7837	Sequence 7837, Ap
45	29.4	8.4	5735	10	US-09-728-952-44	Sequence 44, Appl

## ALIGNMENTS

RESULT 1  
US-09-897-438B-1  
; Sequence 1, Application US/09897438B  
; Patent No. US20020137095A1  
; GENERAL INFORMATION:  
; APPLICANT: Mikoshiba, Katsuhiko  
; APPLICANT: Tate, Ngoko  
; TITLE OF INVENTION: REELIN PROTEIN CR-50 EPTOPE REGION  
; FILE REFERENCE: 04853-0076-00000  
; CURRENT APPLICATION NUMBER: US/09/897,438B  
; CURRENT FILING DATE: 2001-07-03  
; PRIOR APPLICATION NUMBER: JP 2000-202801  
; PRIOR FILING DATE: 2000-07-04  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 351  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-09-897-438B-1

Query Match 100.0%; Score 351; DB 10; Length 351;  
Best Local Similarity 100.0%; Pred. No. 2.6e-106;  
Matches 351; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGCAGTGTGGCACCATCATGCGCAATGCTGTACCTTCTGTGAGCCGTACGGCCCT 60

Db 1 GAGCAGTGTGGCACCATCATGCGCAATGCTGTACCTTCTGTGAGCCGTACGGCCCT 60

QY 61 CGAGAGCTGACCACCATCGCTGAACACACACACATCTGCTCCAGTTTCCATT 120

Db 61 CGAGAGCTGACCACCATCGCTGAACACACACATCTGCTCCAGTTTCCATT 120

QY 121 GGGTCAGGATCATGTCGATTTAGTTACTCTGACCCAGCATCTGTGTATACGCCAAG 180

Db 121 GGGTCAGGATCATGTCGATTTAGTTACTCTGACCCAGCATCTGTGTATACGCCAAG 180

QY 181 AACAAATACCGCTGATTGGATTGAGTTCAGCTGGAGAAAATTTAGACCCCTTCCAATGTGAGCACA 240

Db 181 AACAAATACCGCTGATTGGATTGAGTTCAGCTGGAGAAAATTTAGACCCCTTCCAATGTGAGCACA 240

QY 241 GTATCCACATCTGTACTTCCCGGAGGAGCCCAAGGAGAGCGTGCAGTTCCAGTGG 300

```

Db 241 GTATCCACATCTGTACTCTCCCGAGGAGCCAAAGGGAGGCGTGCAGTCCAGTGG 300
QY 301 AAACAGGACACCTCGCAGTGGGTGAGCTGTATGAGGCGCTCTGGGCCCTG 351
Db 301 AAACAGGACACCTCGCAGTGGGTGAGCTGTATGAGGCGCTCTGGGCCCTG 351

RESULT 2
US-09-880-107-3436
; Sequence 3436, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 3436
; LENGTH: 11580
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 U97916
US-09-880-107-3436

Query Match 75.1%; Score 263.6; DB 10; Length 11580;
Best Local Similarity 84.6%; Pred. No. 9.4e-77;
Matches 296; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 1 GAGCAGTGTGGCAGCATCATGCGTGAACACACACAGCATCTGTCTCCAGTTTCCATT 120
Db 860 GAACAGTGTGGCGGATATGATGCGCAATGCGTCACTCTGTGAACCATATGGCCCA 919
QY 61 CGAGAGCTGACCACCATCGCTGGAACACACACAGCATCTGTCTCCAGTTTCCATT 120
Db 920 CGAAGACTGATTACCAAGCGCTTAATACACACAGCATCTGTCTCCAGTTTCCATT 979
QY 121 GGGTCAGGATCATCTGATTAGTTAGTCTGACCCAGCATCATCTGTCTCCAGTTTCCATT 180
Db 980 GGGTCAGGATCATCTGATTAGTTAGTCTGACCCAGCATCATCTGTCTCCAGTTTCCATT 1039
QY 181 AACAAATACCGCTGATTGGATTTCAGTGGAGAAATAGAGCCCTTCCAAATGTGAGCACA 240
Db 1040 AATAAATCTCGGAGTGAATTCAGTGGAGAAATAGAGCCCTTCCAAATGTGAGCACA 1099
QY 241 GTATCCACATCTGTACTCTCCCGAGGAGCCAAAGGGAGGCGTGCAGTTCAGTGG 300
Db 1100 ATATCCATATCTCTACCTCTCTGAGGAGCCAAAGGGAGGAGTGTCCAAATTCAGTGG 1159
QY 301 AAACAGGACACCTCGCAGTGGGTGAGCTGTATGAGGCGCTCTGGGCCCT 350
Db 1160 AAGCAGAAATCTTCGTGTAGTGAAGTGTATGAAGCCCTGTGGGCCCT 1209

RESULT 3
US-09-864-761-14497/c
; Sequence 14497, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng

```

```

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 14497
; LENGTH: 499
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC000121.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.62
US-09-864-761-14497

Query Match 28.1%; Score 98.8; DB 10; Length 499;
Best Local Similarity 83.6%; Pred. No. 4.4e-23;
Matches 112; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 217 AGAGCCCTTCCAAATGTGAGCAGTCATCCACATCTGTACCTCCCGAGGAGCCAAA 276
Db 487 AGAGCCCTTCCAAATGTGAGCAGTCATCCATATCTCTACCTCTCTGAGGAGCCAAA 428
QY 277 GGGGAGGAGCGTGCAGTTCAGTGGAAACAGGACAGCTCGGAGTGGGTGATGAG 336
Db 427 GGGGAGGAGTGTCCAAATTTTCAGTGGAAAGGAGGAAAATCTTCGTGTAGGTGAAGTGTATGAA 368
QY 337 GCGTCTGGGCCCT 350
Db 367 GCGTCTGGGCCCT 354

RESULT 4
US-09-864-761-31037/c
; Sequence 31037, Application US/09864761
; Patent No. US20020048763A1

```

; GENERAL INFORMATION:  
 ; APPLICANT: Penn, Sharron G.  
 ; APPLICANT: Rank, David R.  
 ; APPLICANT: Hanzel, David K.  
 ; APPLICANT: Chen, Wensheng  
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
 ; FILE REFERENCE: Aecomica-X-1  
 ; CURRENT APPLICATION NUMBER: US/09/864,761  
 ; CURRENT FILING DATE: 2001-05-23  
 ; PRIOR APPLICATION NUMBER: US 60/180,312  
 ; PRIOR FILING DATE: 2000-02-04  
 ; PRIOR APPLICATION NUMBER: US 60/207,456  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: US 09/632,366  
 ; PRIOR FILING DATE: 2000-08-03  
 ; PRIOR APPLICATION NUMBER: GB 24263.6  
 ; PRIOR FILING DATE: 2000-10-04  
 ; PRIOR APPLICATION NUMBER: US 60/236,359  
 ; PRIOR FILING DATE: 2000-09-27  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00664  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00670  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: US 60/234,687  
 ; PRIOR FILING DATE: 2000-09-21  
 ; PRIOR APPLICATION NUMBER: US 09/608,408  
 ; PRIOR FILING DATE: 2000-06-30  
 ; PRIOR APPLICATION NUMBER: US 09/774,203  
 ; PRIOR FILING DATE: 2001-01-29  
 ; NUMBER OF SEQ ID NOS: 49117  
 ; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
 ; SEQ ID NO 31037  
 ; LENGTH: 214  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; OTHER INFORMATION: MAP TO AC000121.1  
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.62  
 ; OTHER INFORMATION: EST\_HUMAN HIT: BE938667.1, EVALUATE 1.00e-76  
 ; OTHER INFORMATION: NT HIT: g14826977, EVALUATE 1.00e-117  
 ; OTHER INFORMATION: SWISSPROT HIT: O14063, EVALUATE 1.20e+00  
 ; US-09-864-761-31037

Query Match 23.2%; Score 81.4; DB 10; Length 214;  
 Best Local Similarity 81.7%; Pred. No. 1.6e-17;  
 Matches 94; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 236 GCACAGTCATCCATCCTGTACCTCCCGAGGAGCCAAAGGAGGAGCGTGCAGTTC 295

Db 214 GCACATCATCCATCCTGTACCTCCCGAGGAGCCAAAGGAGGAGCGTGCAGTTC 155

Qy 296 AGTGAACAGCAGACGACGTCGAGTGGGTGATGAGGCGCTGTGGGCGCCT 350

Db 154 AGTGAACAGCAGAAATCTTCGTGTAGTGAAGTGTATGAGCGCTGTGGGCGCCT 100

RESULT 5  
 US-09-864-761-11808/c  
 ; Sequence 11808, Application US/09864761  
 ; Patent No. US20020048763A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Penn, Sharron G.  
 ; APPLICANT: Rank, David R.  
 ; APPLICANT: Hanzel, David K.  
 ; APPLICANT: Chen, Wensheng  
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
 ; FILE REFERENCE: Aecomica-X-1  
 ; CURRENT APPLICATION NUMBER: US/09/864,761  
 ; CURRENT FILING DATE: 2001-05-23  
 ; PRIOR APPLICATION NUMBER: US 60/180,312  
 ; PRIOR FILING DATE: 2000-02-04  
 ; PRIOR APPLICATION NUMBER: US 60/207,456  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: US 09/632,366  
 ; PRIOR FILING DATE: 2000-08-03  
 ; PRIOR APPLICATION NUMBER: GB 24263.6  
 ; PRIOR FILING DATE: 2000-10-04  
 ; PRIOR APPLICATION NUMBER: US 60/236,359  
 ; PRIOR FILING DATE: 2000-09-27  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00664  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00665  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00670  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: US 60/234,687  
 ; PRIOR FILING DATE: 2000-09-21  
 ; PRIOR APPLICATION NUMBER: US 09/608,408  
 ; PRIOR FILING DATE: 2000-06-30  
 ; PRIOR APPLICATION NUMBER: US 09/774,203  
 ; PRIOR FILING DATE: 2001-01-29  
 ; NUMBER OF SEQ ID NOS: 49117  
 ; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
 ; SEQ ID NO 11808  
 ; LENGTH: 413  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; OTHER INFORMATION: MAP TO AC000121.1  
 ; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.55  
 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.57  
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.51  
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.44  
 ; US-09-864-761-11808

Query Match 15.3%; Score 53.8; DB 10; Length 413;  
 Best Local Similarity 79.0%; Pred. No. 2.8e-08;  
 Matches 64; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 1 GAGCAGTGTGGCACCACATCATGCGCAATGCTGTACCTTCTGTGAGCGCTACGGCCT 60

Db 323 GAACAGTGTGGCGCATATGCGCAATGCGCAATGCGCTTCTGTGAACCATATGGCCCA 264

Qy 61 CGAGAGCTGACCAACACATGC 81



```
Query Match      8.9%; Score 31.4; DB 10; Length 670;
Best Local Similarity 47.3%; Pred. No. 0.85;
Matches 95; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 10 GGCACCATCATGATGCAATGCTGTGCACCTTCTGTGAGCGGTACGGCCCTCGAGAGCTG 69
    |||||
Db 5 GTCTCATCATGATGCTGCTCGTCATCTGTGTGGGGCCCTACTCTTCTCGGTGCTG 64
    |||||

QY 70 ACCACCATGTCCTGAACAACAACAAGCATCTGTCTCCAGCTTTTCCATTGGGTGTCAGGA 129
    |||||
Db 65 CTGGCCGCCGCCAGGAGCCAGACCATGAGCCGCCCTCGCTCCTCAGCGTGGTGGCC 124
    |||||

QY 130 TCATGTCGATTTAGTTACTCTGACCCAGCATCATCTGTGTACGACCAAGACAATACC 189
    |||||
Db 125 GTCTGGCTGACCTGGGCGCAATGGGCGCATCAACCTGTCTATCAGCCATCCGCAATCCC 184
    |||||

QY 190 GCTGATTGGATTGAGCTGGAG 210
    |||||
Db 185 AACATTTCGATGCTCTTAGGG 205
    |||||

RESULT 9
US-10-050-726-1
; Sequence 1, Application us/10050726
; Patent No. US20020151697A1
; GENERAL INFORMATION:
; APPLICANT: Patirana, Marie Sudam
; TITLE OF INVENTION: DNA ENCODING ORPHAN SNORF68 RECEPTOR
; FILE REFERENCE: 60795
; CURRENT APPLICATION NUMBER: US/10/050,726
; CURRENT FILING DATE: 2002-01-16
; PRIOR APPLICATION NUMBER: 09/466,570
; PRIOR FILING DATE: 1999-12-17
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0 - beta
; SEQ ID NO 1
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-050-726-1

Query Match      8.9%; Score 31.4; DB 12; Length 1536;
Best Local Similarity 47.3%; Pred. No. 1.3;
Matches 95; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 10 GGCACCATCATGATGCAATGCTGTGCACCTTCTGTGAGCGGTACGGCCCTCGAGAGCTG 69
    |||||
Db 1022 GTCTCATCATGATGCTGCTCGTCATCTGTGTGGGGCCCTACTGCTTCTCGGTGCTG 1081
    |||||

QY 70 ACCACCATGTCCTGAACAACAACAAGCATCTGTCTCCAGTTTTCATTGGGTGTCAGGA 129
    |||||
Db 1082 CTGGCCGCCGCCAGGAGCCAGACCATGAGCCGCCCTCGCTCCTCAGCGTGGTGGCC 1141
    |||||

QY 130 TCATGTCGATTTAGTTACTCTGACCCAGCATCATCTGTGTACGACCAAGACAATACC 189
    |||||
Db 1142 GTCTGGCTGACCTGGGCGCAATGGGCGCATCAACCTGTCTATCAGCCATCCGCAATCCC 1201
    |||||

QY 190 GCTGATTGGATTGAGCTGGAG 210
    |||||
Db 1202 AACATTTCGATGCTCTTAGGG 1222
    |||||

RESULT 10
US-09-828-644-9
; Sequence 9, Application us/09828644
; Patent No. US20020015998A1
; GENERAL INFORMATION:
; APPLICANT: Vogeli, Gabriel
; TITLE OF INVENTION: No. US20020015998A1el G Protein-Coupled Receptors
; FILE REFERENCE: 00196US1
; CURRENT APPLICATION NUMBER: US/09/828,644
; CURRENT FILING DATE: 2001-04-06
```

```
; PRIOR APPLICATION NUMBER: 60/195,150
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/195,099
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/195,151
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/195,148
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/195,093
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/195,098
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/230,149
; PRIOR FILING DATE: 2000-09-05
; NUMBER OF SEQ ID NOS: 117
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 1860
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-828-644-9

Query Match      8.9%; Score 31.4; DB 10; Length 1860;
Best Local Similarity 47.3%; Pred. No. 1.4;
Matches 95; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 10 GGCACCATCATGATGCAATGCTGTGCACCTTCTGTGAGCGGTACGGCCCTCGAGAGCTG 69
    |||||
Db 1200 GTCTCATCATGATGCTGCTCGTCATCTGTGTGGGGCCCTACTGCTTCTCGGTGCTG 1259
    |||||

QY 70 ACCACCATGTCCTGAACAACAACAAGCATCTGTCTCCAGTTTTCATTGGGTGTCAGGA 129
    |||||
Db 1260 CTGGCCGCCGCCAGGAGCCAGACCATGAGCCGCCCTCGCTCCTCAGCGTGGTGGCC 1319
    |||||

QY 130 TCATGTCGATTTAGTTACTCTGACCCAGCATCATCTGTGTACGACCAAGACAATACC 189
    |||||
Db 1320 GTCTGGCTGACCTGGGCGCAATGGGCGCATCAACCTGTCTATCAGCCATCCGCAATCCC 1379
    |||||

QY 190 GCTGATTGGATTGAGCTGGAG 210
    |||||
Db 1380 AACATTTCGATGCTCTTAGGG 1400
    |||||

RESULT 11
US-09-922-217-605/c
; Sequence 605, Application US/09922217
; Patent No. US20020076414A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Xuqiu
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.471C13
; CURRENT APPLICATION NUMBER: US/09/922,217
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 605
; LENGTH: 417
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-922-217-605
```

```

Query Match      8.9%; Score 31.2; DB 10; Length 417;
Best Local Similarity 50.7%; Pred. No. 0.79;
Matches 75; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 195 TTGATTTCAGCTGAGAAATTAGAGCCCTTCCAAATGTGAGCAGACATCCACATCCT 254
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 339 TTGATTTCAGAGACCCCAATCTAGTGTGGCCAAAGTGCCCACTCTGAGCAAACTCTT 280
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 255 GTACCTCCCGAGGAGCAAGGAGAGCGTGCGAGTTCAGTGGAAACAGGACAGCCT 314
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 279 GGAATGATGGAGAGTCCAGATTACAGAGAAGATCGACGAGAGACACACAGCCA 220
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 315 GCGAGTGGGTGAGTGTATGAGGCTGC 342
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 219 GCAGGCCCCCGAGGCTGAGGCCCTGC 192
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
US-09-833-263-605/c
; Sequence 605, Application US/09833263
; Patent No. US20020110547A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Stoilk, John A.
; APPLICANT: Meagher, Madeleine J.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
; FILE REFERENCE: 210121.471C12
; CURRENT APPLICATION NUMBER: US/09/833,263
; CURRENT FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 1093
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 605
; LENGTH: 417
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-833-263-605

Query Match      8.9%; Score 31.2; DB 10; Length 417;
Best Local Similarity 50.7%; Pred. No. 0.79;
Matches 75; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 195 TTGATTTCAGCTGAGAAATTAGAGCCCTTCCAAATGTGAGCAGACATCCACATCCT 254
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 339 TTGATTTCAGAGACCCCAATCTAGTGTGGCCAAAGTGCCCACTCTGAGCAAACTCTT 280
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 255 GTACCTCCCGAGGAGCAAGGAGAGCGTGCGAGTTCAGTGGAAACAGGACAGCCT 314
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 279 GGAATGATGGAGAGTCCAGAGTTTACAGAGAAGATCGACGAGAGACACACAGCCA 220
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 315 GCGAGTGGGTGAGTGTATGAGGCTGC 342
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 219 GCAGGCCCCCGAGGCTGAGGCCCTGC 192
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
US-10-124-598-4
; Sequence 4, Application US/10124598
; Patent No. US20020119526A1
; GENERAL INFORMATION:
; APPLICANT: Zuker, Charles S.
; APPLICANT: Adler, Jon Elliot
; APPLICANT: Lindemeyer, Juergen
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Nucleic Acids Encoding a G-Protein Coupled Receptor
; FILE REFERENCE: 02307E-088720US
; CURRENT APPLICATION NUMBER: US/10/124,598
; CURRENT FILING DATE: 2002-04-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/361,631
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/112,747

```

```

; PRIOR FILING DATE: EARLIER FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 2532
; TYPE: DNA
; ORGANISM: Mus sp.
; FEATURE:
; OTHER INFORMATION: mouse G-protein coupled receptor (GPCR) B4
; OTHER INFORMATION: nucleotide sequence
US-10-124-598-4

Query Match      8.9%; Score 31.2; DB 12; Length 2532;
Best Local Similarity 47.4%; Pred. No. 1.9;
Matches 93; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 146 ACTCTGACCCCGACATCAGTGTGTATACGCCAAGACAAATACCGTGTATGATTCAGC 205
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 464 ACATTCTCTCTACTTCTCTGTCGCACAGTCCATATAGCGCCATCACCGACAAGCTGC 523
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 206 TGGAGAAATTTAGAGCCCTTCCAAATGTGAGCAGCAGTCATCCACATCTCTTACCTCCCG 265
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 524 AAGACAAGCGCGCTTCCCTGCCATGCTGGCACTGTGCCAGCGGCCACCCACATCG 583
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 266 AGGAAGCCAAAGGGGAGAGCGTCAGTTCAGTGGAAACAGACAGCCTCGAGTGGTG 325
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 584 AGCCATGGTGCAACTGATGTTCACTTCCAGTGGAACTGGATCGTGTGCTGTGAGCG 643
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 326 AGTGTATGAGGCTG 341
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 644 ATGACGATTATGGCCG 659
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
US-09-476-242-22
; Sequence 22, Application US/09476242
; Patent No. US20020146683A1
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: HARTOG, Karin
; APPLICANT: MARTIN, Eric
; TITLE OF INVENTION: MODIFIED HIV ENV POLYPEPTIDES
; FILE REFERENCE: 1605.002
; CURRENT APPLICATION NUMBER: US/09/476,242
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 2298
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Vall120-Ile201;
; OTHER INFORMATION: Ile424-Ala433
US-09-476-242-22

Query Match      8.7%; Score 30.6; DB 10; Length 2298;
Best Local Similarity 52.8%; Pred. No. 2.8;
Matches 66; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 188 CCGCTGATTGGATTCAGCTGGAGAAAATTAGAGCCCTTCCAAATGTGAGCAGTCCATCC 247
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 518 CCGTGTGTAGACCCAGCTGCTGCTGAACGGCAGCCTGCGCAGGAGGCGTGTGTATCC 577
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 248 ACATCTGTACCTCCCGGAGGAGCCAAAGGGGAGAGCGTGCAGTTCAGTGGAAACAGG 307
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 578 GCAGCGAGAAGTTCACCGAAGCGCAAGACCATCATCTGTCAGCTGAAGGAGAGCGTGG 637
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 308 ACAGC 312
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 638 AGATC 642
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 15

US-09-476-242-23  
 ; Sequence 23 Application US/09476242  
 ; Patent No. US20020146683A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: BARNETT, Susan  
 ; APPLICANT: HARTOG, Karin  
 ; APPLICANT: MARTIN, Eric  
 ; TITLE OF INVENTION: MODIFIED HIV ENV POLYPEPTIDES  
 ; FILE REFERENCE: 1605.002  
 ; CURRENT APPLICATION NUMBER: US/09/476,242  
 ; CURRENT FILING DATE: 1999-12-30  
 ; NUMBER OF SEQ ID NOS: 26  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 23  
 ; LENGTH: 2298  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence:  
 ; OTHER INFORMATION: Val120-Ile201B; Ile424-Ala433  
 US-09-476-242-23

Query Match	8.7%	Score 30.6;	DB 10;	Length 2298;
Best Local Similarity	52.8%	Pred. No. 2.8;		
Matches 66;	Conservative 0;	Mismatches 59;	Indels 0;	Gaps 0;

  

QY	188	CCGCTGATTGGATTACAGTGTGAGAAATTTAGAGCCCTTCCAATGTGAGCACAGTCATCC	247
DB	518	CCGTGGTGAGCACCCAGCTGCTGTAACGGCAGCCTGGCGGAGGCGGTGGTGATCC	577
QY	248	ACATCTGTACCTCCCGAGGAGCCAAAGGGGAGAGCGTCCAGTTCAGTGGAAACAGG	307
DB	578	GCACGGAGAACTTCACCGACAACCCCAAGACCACCATCATGTGTCAGCTGAAGGAGAGCGTGG	637
QY	308	ACAGC 312	
DB	638	AGATC 642	

Search completed: November 6, 2002, 16:43:11  
 Job time : 91 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 13:57:21 ; Search time 68 Seconds  
(without alignments)  
1582.993 Million cell updates/sec

Title: US-09-897-438B-1  
Perfect score: 351  
Sequence: 1 gacagtggtggaccatcat.....atgagccctgctgggcccctg 351

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents.NA.\*

1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	351	100.0	11673	4	US-09-334-220-3
2	263.6	75.1	11580	4	US-09-334-220-4
3	35.6	10.1	3344	1	US-07-718-575-13
4	35.6	10.1	3344	1	US-08-481-206-13
5	35.6	10.1	3344	2	US-08-486-269A-13
6	35.6	10.1	3619	4	US-08-377-503-1
7	35.6	10.1	3619	4	US-08-178-019-1
8	31.2	8.9	2532	4	US-09-361-631-4
9	29.8	8.5	2329	4	US-09-411-977-1
10	29.6	8.4	1176	4	US-09-372-422A-25
11	29.6	8.4	5020	3	US-08-938-291A-3
12	29.4	8.4	2657	4	US-09-488-671-3
13	29	8.3	1011	4	US-09-641-806-3
14	29	8.3	1026	4	US-09-641-806-1
15	29	8.3	2095	1	US-08-336-408B-3
16	29	8.3	2095	5	PCT-US91-00399-3
17	28.8	8.2	2565	4	US-08-985-492-7
18	28.6	8.1	935	4	US-09-221-017B-331
19	28.6	8.1	1068	5	PCT-US91-00899-13
20	28.6	8.1	1215	5	PCT-US91-00899-5
21	28.6	8.1	1488	1	US-07-914-281-9
22	28.6	8.1	1488	1	US-08-393-246-9
23	28.6	8.1	1488	1	US-08-525-058A-9
24	28.6	8.1	1488	2	US-08-696-731-9
25	28.6	8.1	1488	4	US-09-042-531-9
26	28.6	8.1	2134	2	US-08-483-151-3
27	28.6	8.1	2134	5	PCT-US96-06427-3

c 28	28.6	8.1	2175	4	US-08-482-073-9	Sequence 9, Appli
c 29	28.6	8.1	2861	4	US-08-482-073-10	Sequence 10, Appli
c 30	28.6	8.1	3647	1	US-07-914-281-7	Sequence 7, Appli
c 31	28.6	8.1	3647	1	US-08-393-246-7	Sequence 7, Appli
c 32	28.6	8.1	3647	1	US-08-525-058A-7	Sequence 7, Appli
c 33	28.6	8.1	3647	2	US-08-696-731-7	Sequence 7, Appli
c 34	28.6	8.1	3647	4	US-09-042-531-7	Sequence 7, Appli
c 35	28.6	8.1	3647	5	PCT-US91-00899-4	Sequence 4, Appli
c 36	28	8.0	1185	3	US-09-023-339-3	Sequence 3, Appli
c 37	28	8.0	1260	3	US-09-023-173-5	Sequence 5, Appli
c 38	28	8.0	1308	3	US-09-023-173-10	Sequence 10, Appli
c 39	28	8.0	1308	3	US-09-023-339-6	Sequence 6, Appli
c 40	27.6	7.9	1456	4	US-09-056-285A-6	Sequence 6, Appli
c 41	27.6	7.9	1473	4	US-09-056-285A-9	Sequence 9, Appli
c 42	27.6	7.9	1605	4	US-09-149-476-187	Sequence 187, App
c 43	27.6	7.9	1871	1	US-08-416-478A-1	Sequence 1, Appli
c 44	27.6	7.9	1871	2	US-08-474-988B-1	Sequence 1, Appli
c 45	27.6	7.9	1871	2	US-08-394-442B-1	Sequence 1, Appli

#### ALIGNMENTS

RESULT 1  
US-09-334-220-3  
; Sequence 3, Application US/09334220  
; Patent No. 6323177  
; GENERAL INFORMATION:  
; APPLICANT: St. Jude's Children's Research Hospital  
; APPLICANT: Curran, Thomas  
; APPLICANT: D'Arcangelo, Gabriella  
; TITLE OF INVENTION: INTERACTION OF REELIN WITH VERY LOW  
; TITLE OF INVENTION: DENSITY LIPOPROTEIN (VLDL) RECEPTOR FOR SCREENING AND  
; TITLE OF INVENTION: THERAPIES  
; FILE REFERENCE: 2427/06704  
; CURRENT APPLICATION NUMBER: US/09/334,220  
; CURRENT FILING DATE: 1999-06-16  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FASTSEQ for Windows Version 3.0  
; SEQ ID NO 3  
; LENGTH: 11673  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-09-334-220-3

Query Match	100.0%	Score	351	DB	4	Length	11673
Best Local Similarity	100.0%	Pred. No.	1.7e-108				
Matches	351	Conservative	0	Mismatches	0	Indels	0
Gaps	0						
Qy	1	GAGCAGTGTGGCACCACCATCATGCGCAATGCTGTACCTTCTGTGAGCCGTACGGCCCT	60				
Db	970	GAGCAGTGTGGCACCACCATCATGCGCAATGCTGTACCTTCTGTGAGCCGTACGGCCCT	1029				
Qy	61	CGAGAGTGTGACCCACCATGCTGAACACACACAGCATCTGCTCCCTCCAGTTTTCATT	120				
Db	1030	CGAGAGTGTGACCCACCATGCTGAACACACACAGCATCTGCTCCCTCCAGTTTTCATT	1089				
Qy	121	GGGTGAGATCATGTGCAATTTAGTTACTTCTGACCCAGCATCTGTGTATACGCCAAG	180				
Db	1090	GGGTGAGATCATGTGCAATTTAGTTACTTCTGACCCAGCATCTGTGTATACGCCAAG	1149				
Qy	181	AACATACCGTGTGATTCAGTTCAGTGGAGAAATAGAGCCCTTCCAATGTGAGCACA	240				
Db	1150	AACATACCGTGTGATTCAGTTCAGTGGAGAAATAGAGCCCTTCCAATGTGAGCACA	1209				
Qy	241	GTATTCACATCTCTACCTCCCGAGGAGCAAGCCAAAGGAGAGCGTGCAGTTCCAGTGG	300				
Db	1210	GTATTCACATCTCTACCTCCCGAGGAGCAAGCCAAAGGAGAGCGTGCAGTTCCAGTGG	1269				
Qy	301	AAACAGGACACGCTCGGAGTGGGTGAGTGATGAGGCCCTGTGGGCCCTG	351				
Db	1270	AAACAGGACACGCTCGGAGTGGGTGAGTGATGAGGCCCTGTGGGCCCTG	1320				

RESULT 2  
US-09-334-220-4  
; Sequence 4, Application US/09334220  
; Patent No. 6323177  
; GENERAL INFORMATION:  
; APPLICANT: St. Jude Children's Research Hospital  
; APPLICANT: Curran, Thomas  
; APPLICANT: D'Arcangelo, Gabriella  
; TITLE OF INVENTION: INTERACTION OF REELIN WITH VERY LOW  
; TITLE OF INVENTION: DENSITY LIPOPROTEIN (VLDL) RECEPTOR FOR SCREENING AND  
; TITLE OF INVENTION: THERAPIES  
; FILE REFERENCE: 2427/07704  
; CURRENT APPLICATION NUMBER: US/09/334,220  
; CURRENT FILING DATE: 1999-06-16  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 11580  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-334-220-4

Query Match 75.1%; Score 263.6; DB 4; Length 11580;  
Best Local Similarity 84.6%; Pred. No. 6.2e-79;  
Matches 296; Conservative 0; Mismatches 54; Indels 0; Gaps 0;  
Qy 1 GAGCAGTGTGGCAGCATCATGCGCAATGCTGTACCTTCTGTGAGCGGTACGGCCT 60  
Db 860 GAACAGTGTGGCGGATGATGATGCGCAATGCGCTCCTTCTGTGAGCAATATGGCCCA 919  
Qy 61 CGAGAGTGTGACACACATGCTGAAACACACACACATGTCCTCCAGTTTCCATT 120  
Db 920 CGAGAACTGATTACACAGCGCTTAAACACACACAGCTTCTGCTCCCAATTTCCATT 979  
Qy 121 GGTTCAGATCATGCTGATTTAGTTACTCTGACCCAGCATCATGTCATACGCCAAG 180  
Db 980 GGGTCAGTTCATGCTGCTTTAGTTATTCAGACCCAGCATCATGTTATATATGCCAAG 1039  
Qy 181 AACATACCGCTGATGATTGATTCAGCTGGAGAAATTTAGAGCCCTTCCAAATGTGAGCACA 240  
Db 1040 AATACTCTGGGACTGGATTCAGTACAGAGAAATTTAGAGCCCTTCCAAATGTGAGCACA 1099  
Qy 241 GTCATCCACATCTGTACCTCCCGAGAGAGCCAAAGAGAGCGGTGCGAGTTCCAGTGG 300  
Db 1100 ATCATCCATATCTCTACCTTCTCTGAGGAGCGCCAAAGGGAGAAATGTCCAAATTTCACTGG 1159  
Qy 301 AACAGGACAGCTGCGAGTGGGTGAGTGTATGAGGCGCTGCTGGGCTCT 350  
Db 1160 AAGCAGGAAATCTTCGTGTAGGTGAAGTGTATGAGGCGCTGCTGGGCTCT 1209

RESULT 3  
US-07-718-575-13  
; Sequence 13, Application US/07718575  
; Patent No. 5202257  
; GENERAL INFORMATION:  
; APPLICANT: Heinemann Ph.D., Stephen F.  
; APPLICANT: Boulter Ph.D., James R.  
; APPLICANT: Hollmann Ph.D., Michael NMN  
; APPLICANT: Bettler Ph.D., Bernhard NMN  
; APPLICANT: Jensen Ph.D., Jan E.  
; TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS AND  
; TITLE OF INVENTION: METHODS  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark  
; STREET: 444 So. Flower St., Suite 2000  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: United States  
; ZIP: 90071-2921  
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA: US/07/718,575  
; APPLICATION NUMBER: US/07/718,575  
; FILING DATE: 19910813  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Reiter Ph.D., Stephen E.  
; REGISTRATION NUMBER: 31192  
; REFERENCE/DOCKET NUMBER: P31 8962  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; TELEX: 9103330318  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3344 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; ORIGINAL SOURCE:  
; CELL TYPE:  
; CELL LINE:  
; IMMEDIATE SOURCE:  
; LIBRARY: Glur7  
; CLONE: Glur7  
; NAME/KEY: CDS  
; LOCATION: 1..2766  
US-07-718-575-13  
Query Match 10.1%; Score 35.6; DB 1; Length 3344;  
Best Local Similarity 53.6%; Pred. No. 0.042;  
Matches 74; Conservative 0; Mismatches 64; Indels 0; Gaps 0;  
Qy 214 ATTAGAGCCCTTCCAAATGTGAGCAGCATCATCCACATCTGTACCTCCCGAGGAAGCC 273  
Db 2317 ATGGCTTCCCTTACAGGACAAATCACCATCCCATTTCTGACGTTCAGGAGGAGAC 2376  
Qy 274 AAGGGGAGAGCGGTGAGTTCCAGTGGAAACAGGACAGCCTGCGAGTGGGTGAGGTGTAT 333  
Db 2377 AAGCTGCACATCATGAAGAGAGTGTGGGAGGAGCGGGTCCCGGAGGAGGAAC 2436  
Qy 334 GAGCCTGCTGGGCTTG 351  
Db 2437 AAGAGCCAGCGCACTG 2454  
RESULT 4  
US-08-481-206-13  
; Sequence 13, Application US/08481206  
; Patent No. 5739291  
; GENERAL INFORMATION:  
; APPLICANT: Heinemann Ph.D., Stephen F.  
; APPLICANT: Boulter Ph.D., James R.  
; APPLICANT: Hollmann Ph.D., Michael NMN  
; APPLICANT: Bettler Ph.D., Bernhard NMN  
; APPLICANT: Jensen Ph.D., Jan E.  
; TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS AND  
; TITLE OF INVENTION: METHODS  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark  
; STREET: 444 So. Flower St., Suite 2000  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: United States  
; ZIP: 90071-2921  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk

```
;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/481,206
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/013,767
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter Ph.D., Stephen E.
; REGISTRATION NUMBER: 31192
; REFERENCE/DOCKET NUMBER: P31 8962
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEX: 9103330318
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3344 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; ORIGINAL SOURCE:
; CELL TYPE:
; IMMEDIATE SOURCE:
; LIBRARY:
; CLONE: Glu7
; NAME/KEY: CDS
; LOCATION: 1..2766
;
US-08-481-206-13

Query Match 10.1%; Score 35.6; DB 1; Length 3344;
Best Local Similarity 53.6%; Pred. No. 0.042;
Matches 74; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 214 ATTAGAGCCCTTCCATGTGAGCAGTATCCATCCATCTGTACCTCCCGAGGAAGCC 273
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2317 ATGGCTCCCTTACAGGGACAAAATCACCATCGCCATTTCTGCAGCTGCAGGAGGAGC 2376
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 274 AAAGGGAGAGCGTCCAGTTCAGTGGAAACAGACAGCAGCCTGCGAGTGGTGAGGTGTAT 333
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2377 AAGCTGCACATCATGAAGGAGAGTGGTGGCGAGCAGCGGGTGCCCGGAGGAGGAGAAC 2436
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 334 GAGGCCTGCTGGGCCCTG 351
   ||||| ||||| |||||
Db 2437 AAGGAGGCCAGCGCACTG 2454

RESULT 5
US-08-486-269A-13
; Sequence 13, Application US/08486269A
; Patent No. 5945509
; GENERAL INFORMATION:
; APPLICANT: Heinemann, Stephen F.
; APPLICANT: Boulter, James R.
; APPLICANT: Hollmann, Michael
; APPLICANT: Bettler, Bernhard
; APPLICANT: Jensen, Jan E.
; TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gray Cary Ware & Freidenrich LLP
; STREET: 4365 Executive Drive, Suite 1600
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
```

```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows DEMONSTRATION Version 2.0D
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,269A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/013,767
; FILING DATE: 04-FEB-1993
; APPLICATION NUMBER: 07/718,575
; FILING DATE: 21-JUN-1991
; APPLICATION NUMBER: PCT/US90/06153
; FILING DATE: 25-OCT-1990
; APPLICATION NUMBER: 07/428,116
; FILING DATE: 27-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9986
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-677-1409
; TELEX: 619-677-1465
;
INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3344 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; IMMEDIATE SOURCE:
; CLONE: Glu7
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1..2763
; OTHER INFORMATION:
;
US-08-486-269A-13

Query Match 10.1%; Score 35.6; DB 2; Length 3344;
Best Local Similarity 53.6%; Pred. No. 0.042;
Matches 74; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 214 ATTAGAGCCCTTCCATGTGAGCAGTATCCATCCATCTGTACCTCCCGAGGAAGCC 273
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2317 ATGGCTCCCTTACAGGGACAAAATCACCATCGCCATTTCTGCAGCTGCAGGAGGAGC 2376
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 274 AAAGGGAGAGCGTCCAGTTCAGTGGAAACAGACAGCAGCCTGCGAGTGGTGAGGTGTAT 333
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2377 AAGCTGCACATCATGAAGGAGAGTGGTGGCGAGCAGCGGGTGCCCGGAGGAGGAGAAC 2436
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 334 GAGGCCTGCTGGGCCCTG 351
   ||||| ||||| |||||
Db 2437 AAGGAGGCCAGCGCACTG 2454

RESULT 6
US-08-377-503-1
; Sequence 1, Application US/08377503
; Patent No. 6322999
; GENERAL INFORMATION:
; APPLICANT: Kamboj, Rajender
; APPLICANT: Elliot, Candace E.
; APPLICANT: Nutt, Stephen L.
; TITLE OF INVENTION: Kainate-Binding, Human CNS Receptors of
; TITLE OF INVENTION: the EAA5 Family
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: Suite 500, 3000 K Street
; CITY: N.W.
```

```

STATE: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Review #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/377,503
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/945,210
FILING DATE: 17-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Bent, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 1677/192
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3619 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 16..108
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 109..2772
FEATURE:
NAME/KEY: CDS
LOCATION: 16..2772
US-08-377-503-1

```

	Query Match	10.1%	Score 35.6;	DB 4;	Length 3619;
	Best Local Similarity	53.6%;	Pred. No. 0.043;		
	Matches	74;	Conservative	0;	Mismatches 64; Indels 0; Gaps 0;
QY	214	ATTAGAGCCCCCTTCCATCTGAGCACAGTATATCCACATCTCTGTACCTCCCGAGGAGGCC	273		
Db	2326	ATGGGCTCCCCATACCGGGACAAGATCACCATGCCATCCCTGTGCAGTCTCAGGAGGAGGAC	2385		
QY	274	AAAGGGGAGAGCGCTGCAGTTCCTCAGTGGAAACAGSACAGCCTGCCAGTGGGGTGAGGTGTAT	333		
Db	2386	AACTGCATATCATGAGGAGAGTGTGGGGGGCAGCGGTGCTCTGAGGAGGAGAAAC	2445		
QY	334	GAGGCTGCTGGGCCCTG	351		
Db	2446	AAAGAGGCGAGTCCCTG	2463		

RESULT 7  
 US-08-178-019-1  
 ; Sequence 1, Application US/08178019  
 ; Patent No. 6333161  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kamboj, Rajender  
 ; APPLICANT: Elliot, Candace E.  
 ; APPLICANT: Nutt, Stephen L.  
 ; TITLE OF INVENTION: Kainate-Binding, Human CNS Receptors of  
 ; TITLE OF INVENTION: the EAA5 Family  
 ; NUMBER OF SEQUENCES: 13  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Foley & Lardner  
 ; STREET: Suite 500, 3000 K Street  
 ; CITY: N.W.

3

STATE: Washington, D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/178,019  
FILING DATE: 06-JAN-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Bent, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 1677/227  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3619 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 16..108  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 109..2772  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 16..2772  
US-08-178-019-1

[illegible]

```

RESULT 8
US-09-361-631-4
; Sequence 4, Application US/09361631
; Patent No. 6383778
; GENERAL INFORMATION:
; APPLICANT: Zuker, Charles S.
; APPLICANT: Adler, Jon Elliot
; APPLICANT: Lindemeier, Juergen
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Nucleic Acids Encoding a G-Protein Coupled Receptor
; TITLE OF INVENTION: Involved in Sensory Transduction
; FILE REFERENCE: 02307E-0887200S
; CURRENT APPLICATION NUMBER: US/09/361,631
; CURRENT FILING DATE: 1999-07-27
; EARLIER APPLICATION NUMBER: US 60/095,464
; EARLIER FILING DATE: 1998-07-28
; EARLIER APPLICATION NUMBER: US 60/112,747
; EARLIER FILING DATE: 1998-12-17

```

32



OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FastSeq  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/938, 291A  
FILING DATE: September 26, 1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/027,337  
FILING DATE: October 11, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 228/172  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5020 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-938-291A-3

Query Match 8.4%; Score 29.6; DB 3; Length 5020;  
Best Local Similarity 56.0%; Pred. No. 5.4;  
Matches 56; Conservative 0; Mismatches 44; Indels 0; Gaps 0;  
QY 238 ACAGTCATCCATCCCTGACCTCCCGGAGGAGCCAAAGGAGCGGTGCAGTTCAG 297  
Db 984 ACAGTGTCCATCATCTCCCAAGTCTGCTCGGTCTCCAAAGGAGCGGTGCAGTTC 1043  
QY 298 TGAAGACAGGACACCTCGGAGTGGGTGAGGTGTATGAGG 337  
Db 1044 CCACAGCATCTCAGAGTGGAGATGCAGAGTATGGCCAGG 1083

RESULT 12  
US-09-488-671-3  
Sequence 3, Application US/09488671A  
Patent No. 6187545  
GENERAL INFORMATION:  
APPLICANT: Robert McKay  
APPLICANT: Madeline M. Butler  
APPLICANT: Jacqueline Wyatt  
APPLICANT: Lex M. Cowser  
TITLE OF INVENTION: ANTISENSE MODULATION OF PEPCK-CYTOSOLIC EXPRESSION  
FILE REFERENCE: RTS-0123  
CURRENT APPLICATION NUMBER: US/09/488, 671A  
CURRENT FILING DATE: 2000-01-21  
NUMBER OF SEQ ID NOS: 177  
SEQ ID NO 3  
LENGTH: 2657  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (122)...(1990)  
US-09-488-671-3

Query Match 8.4%; Score 29.4; DB 4; Length 2657;  
Best Local Similarity 55.3%; Pred. No. 4.6;  
Matches 57; Conservative 0; Mismatches 46; Indels 0; Gaps 0;  
QY 243 CATCCACATCTCTACTCCCGGAGGAGCCAAAGGAGCGGTGCAGTTCAGTTGGAA 302  
Db 247 CATCCACATCTGTGACGGCTCTCAGGAGGAGAAATGGCGGCTTCTGGGCCAGATGGAGGA 306  
QY 303 ACAGGACAGCCTCGGAGTGGGTGAGGTGTATGAGGCTGTCTGG 345  
Db 307 AGAGGCGATCCTCAGCGCGTGAAGAAGTATGACAACTGCTGTG 349

RESULT 13  
US-09-641-806-3  
Sequence 3, Application US/09641806  
Patent No. 6395527  
GENERAL INFORMATION:  
APPLICANT: Beraud, Christophe  
APPLICANT: Freedman, Richard  
TITLE OF INVENTION: No. 6395527el motor proteins and methods for  
TITLE OF INVENTION: their use  
FILE REFERENCE: 1034  
CURRENT APPLICATION NUMBER: US/09/641,806  
CURRENT FILING DATE: 2000-08-17  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 1011  
TYPE: DNA  
ORGANISM: Human  
US-09-641-806-3

Query Match 8.3%; Score 29; DB 4; Length 1011;  
Best Local Similarity 54.1%; Pred. No. 3.9;  
Matches 59; Conservative 0; Mismatches 50; Indels 0; Gaps 0;  
QY 207 GGAGAAATTTAGAGCCCTTCCAAATGTGAGCAGATCATCCACATCTGTACCTCCCGGA 266  
Db 411 GGAGGAGTTCCGAGACCTGCTCGAGGTGGGCACTGCCAGCGGTGACATCCAGCTCCGGGA 470  
QY 267 GGAAGCCAAAGGGAGAGCGGTGCAGTTCAGTGGAAACAGACAGGCGTG 315  
Db 471 AGATGAGCGCGGAATGTTGTGCTGTGCGGGGTGAAGGAGGTGCAGCGT 519

RESULT 14  
US-09-641-806-1  
Sequence 1, Application US/09641806  
Patent No. 6395527  
GENERAL INFORMATION:  
APPLICANT: Beraud, Christophe  
APPLICANT: Freedman, Richard  
TITLE OF INVENTION: No. 6395527el motor proteins and methods for  
TITLE OF INVENTION: their use  
FILE REFERENCE: 1034  
CURRENT APPLICATION NUMBER: US/09/641,806  
CURRENT FILING DATE: 2000-08-17  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 1026  
TYPE: DNA  
ORGANISM: Human  
US-09-641-806-1

Query Match 8.3%; Score 29; DB 4; Length 1026;  
Best Local Similarity 54.1%; Pred. No. 4;  
Matches 59; Conservative 0; Mismatches 50; Indels 0; Gaps 0;  
QY 207 GGAGAAATTTAGAGCCCTTCCAAATGTGAGCAGATCATCCACATCTGTACCTCCCGGA 266  
Db 414 GGAGGAGTTCCGAGACCTGCTCGAGGTGGGCACTGCCAGCGGTGACATCCAGCTCCGGGA 473  
QY 267 GGAAGCCAAAGGGAGAGCGGTGCAGTTCAGTGGAAACAGACAGGCGTG 315  
Db 474 AGATGAGCGCGGAATGTTGTGCTGTGCGGGGTGAAGGAGGTGCAGGTG 522

RESULT 15  
US-08-336-408B-3/c  
Sequence 3, Application US/08336408B  
Patent No. 5723329  
GENERAL INFORMATION:  
APPLICANT: EVANS, RONALD M.

APPLICANT: MANGELSDORF, DAVID J.  
TITLE OF INVENTION: RETINOID RECEPTOR COMPOSITIONS AND  
METHODS  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PRETTY, SCHROEDER, BRUEGEMANN & CLARK  
STREET: 444 SOUTH FLOWER STREET, SUITE 2000  
CITY: LOS ANGELES  
STATE: CA  
COUNTRY: USA  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/336,408B  
FILING DATE: 08-NOV-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/933,453  
FILING DATE: 21-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US91/00399  
FILING DATE: 22-JAN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/478,071  
FILING DATE: 09-FEB-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: REITER, STEPHEN E.  
REGISTRATION NUMBER: 31,192  
REFERENCE/DOCKET NUMBER: P41 9851  
TELEPHONE: 619-546-1995  
TELEFAX: 619-546-9392  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2095 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: both  
MOLECULE TYPE: cdna  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 178..1581  
US-08-336-408B-3

Query Match 8.3%; Score 29; DB 1; Length 2095;  
Best Local Similarity 53.0%; Pred. No. 5.6;  
Matches 62; Conservative 0; Mismatches 55; Indels 0; Gaps 0;  
QY 14 CCATCATGATGCGATGCTGTACCTTCTGTAGCGGTACGGCCCTCGAGAGCTGACCA 73  
Db 1438 CAAACCTGCCCGCTGCTCAGGGTACTTGTGTTGACGTACGCTTCTAGTGACGCATACA 1379  
QY 74 CCACATGCTCTGAACACAGCATCTCTCCATTTCCATTGGGTCCAGGT 130  
Db 1378 CCTTCCTCAACGGCTCCACCTCAGAGGGTTTGAGAGCCCTTAGAGTCAGGGT 1322

Search completed: November 6, 2002, 15:28:27  
Job time : 81 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 11:43:51 : Search time 300 Seconds  
(without alignments)  
2634.838 Million cell updates/sec

Title: US-09-897-438B-1  
Perfect score: 351  
Sequence: 1 gacgagtgtgaccatcat.....atgagcctctgggcctg 351

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_101002:\*

- 1: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*
- 2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*
- 3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*
- 4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*
- 5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*
- 6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*
- 7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*
- 8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*
- 9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*
- 10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*
- 11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*
- 12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*
- 13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*
- 14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*
- 15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*
- 16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*
- 17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*
- 18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*
- 19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*
- 20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*
- 21: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*
- 22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:\*
- 23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*
- 24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	351	100.0	351	24	ABL40165
2	351	100.0	2745	24	AAD22754
3	351	100.0	11673	24	ABA92603
4	351	100.0	11673	24	AB199284
5	263.6	75.1	11580	24	ABN96939
6	263.6	75.1	11580	24	ABA92604
7	263.6	75.1	11632	23	AAS9484
8	195.8	55.8	2025	24	AAD22779
9	195.8	55.8	2274	24	AAD22753
					Mouse reelin prote
					Mus musculus trunc
					Mouse reelin encod
					Mouse ischaemic co
					Gene #3437 used to
					Human reelin encod
					DNA encoding novel
					Xenopus laevis int
					Xenopus laevis tru

c	10	53.8	15.3	413	22	AAK07720	Human brain expres
c	11	53.8	15.3	413	22	AAK33553	Human bone marrow
c	12	53.8	15.3	413	24	ABS08422	Human genome-deriv
c	13	46.2	13.2	87	22	AAK20384	Human brain expres
c	14	46.2	13.2	87	22	AAK46490	Human bone marrow
c	15	46.2	13.2	87	24	ABS20836	Human genome-deriv
c	16	36.4	10.4	5994	23	ABLI0961	Drosophila melanog
c	17	36.4	10.4	8767	23	ABLI0960	Drosophila melanog
c	18	35.6	10.1	3619	15	AAQ44685	Excitatory amino a
c	19	32	9.1	1473	24	ABL40040	Synthetic Env gp12
c	20	32	9.1	1986	24	ABL40041	Synthetic Env gp14
c	21	32	9.1	1986	24	ABL40042	Synthetic Env gp14
c	22	32	9.1	2052	24	ABL40050	Synthetic Env gp14
c	23	32	9.1	2073	24	ABL40051	Synthetic Env gp14
c	24	32	9.1	2073	24	ABL40052	Synthetic gp140 po
c	25	32	9.1	2397	24	ABL40043	Synthetic Env gp16
c	26	32	9.1	2529	24	ABL40044	Synthetic Env gp16
c	27	32	9.1	2529	24	ABL40045	Synthetic Env gp16
c	28	32	9.1	2610	24	ABL39970	Synthetic Env gp16
c	29	32	9.1	2613	24	ABL40046	Synthetic Env gp16
c	30	32	9.1	2616	24	ABL40047	Synthetic Env gp16
c	31	32	9.1	2616	24	ABL40048	Synthetic Env gp16
c	32	31.8	9.1	1759	23	ABLO1883	Drosophila melanog
c	33	31.8	9.1	4338	23	ABLO1882	Drosophila melanog
c	34	31.6	9.0	5332	23	ABLO3708	Drosophila melanog
c	35	31.6	9.0	9496	22	AAO2854	Human reproductive
c	36	31.6	9.0	9496	23	ABL97329	Human testicular a
c	37	31.4	8.9	670	24	AAS62851	Human G-protein-co
c	38	31.4	8.9	1550	22	AAI71602	Human G-protein co
c	39	31.4	8.9	1604	24	AAD36738	Human G-protein-co
c	40	31.4	8.9	1860	24	AAS62852	Human G-protein-co
c	41	31.4	8.9	2177	24	AAD24965	Human G-protein co
c	42	31.4	8.9	5521	22	AAI71603	Human G-protein co
c	43	31.2	8.9	417	22	AAI29056	Colon tumour relat
c	44	31.2	8.9	2203	22	AAK51633	Human polynucleoti
c	45	31.2	8.9	2532	21	AAZ58964	Mouse GPCR-B4 poly

## ALIGNMENTS

## RESULT 1

ABL40165  
ID ABL40165 standard; DNA; 351 BP.

AC ABL40165;

DT 21-MAY-2002 (first entry)

DE Mouse reelin protein CR-50 epitope region encoding DNA SEQ ID NO:1.

KW Mouse; reelin protein CR-50 epitope region; elucidation; neuron;  
cerebral disturbance; reelin protein; neuroprotective; gene; ds.

OS Mus musculus.

FH Key Location/Qualifiers

FT CDS 1..351

FT /tag= a

FT /partial

FT /product= "reelin protein CR-50 epitope region"

FT /note= "no start or stop codons given"

PN JP2002017361-A.

XX 22-JAN-2002.

XX 04-JUL-2000; 2000JP-0202801.

XX 04-JUL-2000; 2000JP-0202801.

XX (RIKA ) RIKAGAKU KENKYUSHO.

DR WPI; 2002-221707/28.  
XX P-PSDB; ABB06244.  
PT Reelin protein CR-50 epitope region, useful for diagnosis and treatment  
PT of cerebral disturbance  
XX Claim 5; Page 11; 16pp; Japanese.  
XX The present sequence encodes the mouse reelin protein CR-50 epitope  
CC region, which contains the CR-50 antibody recognition site and is free  
CC from F-spondin domains and repetitive sites. Also described are: (1) an  
CC expression vector comprising a polynucleotide encoding a reelin protein  
CC epitope region; (2) host cells with transfected the expression vector;  
CC (3) polypeptides prepared by culture of the host cells; and (4)  
CC polynucleotides comprising the 351 base sequence given in ABL40165 which  
CC encodes the 117 amino acid sequence given in ABB06244; and (5) use of  
CC the polynucleotide for diagnosis and/or treatment of diseases caused by  
CC abnormal positioning of neural cells, and stimulation of association of  
CC reelin protein. The mouse reelin protein CR-50 epitope region has  
CC neuroprotective activity, and can be used in the diagnosis and treatment  
CC of cerebral disturbance due to an abnormal reelin gene and positioning  
XX of neurons.  
XX Sequence 351 BP; 86 A; 98 C; 92 G; 75 T; 0 other;  
XX  
XX Query Match 100.0%; Score 351; DB 24; Length 351;  
XX Best Local Similarity 100.0%; Pred. No. 3.7e-103;  
XX Matches 351; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GAGCAGTGTGGCACCACATCATCGGCAATGCTGTACCTTCTGTGAGCCGTACGGCCCT 60  
DB 1 GAGCAGTGTGGCACCACATCATCGGCAATGCTGTACCTTCTGTGAGCCGTACGGCCCT 60  
QY 61 CGAGAGTGTGGCACCACATCGCTGAACACACACAGCATCTGTCTCCAGTTTCCATT 120  
DB 61 CGAGAGTGTGGCACCACATCGCTGAACACACACAGCATCTGTCTCCAGTTTCCATT 120  
QY 121 GGGTCAGGATCATGTCGATTAGTTACTTCTGACCCAGCATCTGTCTATAGCCCAAG 180  
DB 121 GGGTCAGGATCATGTCGATTAGTTACTTCTGACCCAGCATCTGTCTATAGCCCAAG 180  
QY 181 AACAAATACCGTGTGATTGATTCAGTGGAGAAATAGAGCCCTTCCAATGTGAGCACA 240  
DB 181 AACAAATACCGTGTGATTGATTCAGTGGAGAAATAGAGCCCTTCCAATGTGAGCACA 240  
QY 241 GTCATCCACATCTGTACTCTCCGAGGAGCCAAAGAGGAGCGTGCAGTTCCAGTGG 300  
DB 241 GTCATCCACATCTGTACTCTCCGAGGAGCCAAAGAGGAGCGTGCAGTTCCAGTGG 300  
QY 301 AAACAGGACAGCCCTGCGAGTGGGTGAGGTGTATGAGGCCCTGCTGGGCCCTG 351  
DB 301 AAACAGGACAGCCCTGCGAGTGGGTGAGGTGTATGAGGCCCTGCTGGGCCCTG 351  
RESULT 2  
AAD22754  
ID AAD22754 standard; cDNA; 2745 BP.  
XX  
XX AAD22754;  
XX  
XX 26-FEB-2002 (first entry)  
XX  
XX Mus musculus truncated reelin cDNA.  
XX  
XX Mouse; reelin; F-spondin domain; CR-50 epitope; gene therapy; agyria;  
XX polymicrogyria; ectopic gray matter; ss.  
XX  
XX Mus musculus.  
XX  
XX Key Location/Qualifiers  
XX CD5 283..2052  
XX /\*tag= a  
XX /product= "Mouse truncated reelin protein"

FT sig\_peptide 283..363  
FT /\*tag= b  
FT mat\_peptide 364..2049  
FT /\*tag= c  
FT /product= "Mature truncated reelin protein"  
FT misc\_feature 284..849  
FT /\*tag= d  
FT /note= "Encodes F-spondin domain"  
FT 970..1320  
FT /\*tag= e  
FT /note= "Encodes CR-50 epitope region"  
XX  
XX EPI149844-A2.  
XX 31-OCT-2001.  
XX 11-APR-2001; 2001EP-0303411.  
XX 11-APR-2000; 2000JP-0109954.  
XX (RIKE ) RIKEN KK.  
XX Mikoshiba K, Tabata H, Nakajima K;  
XX WPI; 2002-019320/03.  
XX P-PSDB; AAE13606.  
XX Novel truncated Reelin protein containing F-spondin domain and CR-50  
XX recognition site of Reelin protein, but not having Reelin repeat site,  
XX useful to treat diseases including agyria due to abnormal neuron  
XX alignment  
XX  
XX Claim 10; Page 20-26; 47pp; English.  
XX  
XX The invention relates to a truncated Reelin protein comprising a  
XX F-spondin domain and a CR-50 recognition site but no reelin protein  
XX laminated structure of cerebrum. The truncated reelin protein and its  
XX DNA are useful for treating diseases including agyria, polymicrogyria,  
XX and ectopic gray matter due to abnormal neuronal alignment. Truncated  
XX reelin protein DNA is useful in gene therapy. The present sequence is  
XX a cDNA encoding Mus musculus truncated reelin protein.  
XX  
XX Sequence 2745 BP; 661 A; 716 C; 714 G; 654 T; 0 other;  
XX  
XX Query Match 100.0%; Score 351; DB 24; Length 2745;  
XX Best Local Similarity 100.0%; Pred. No. 9e-103;  
XX Matches 351; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GAGCAGTGTGGCACCACATCATCGGCAATGCTGTACCTTCTGTGAGCCGTACGGCCCT 60  
DB 970 GAGCAGTGTGGCACCACATCATCGGCAATGCTGTACCTTCTGTGAGCCGTACGGCCCT 1029  
QY 61 CGAGAGTGTGGCACCACATCGCTGAACACACACAGCATCTGTCTCCAGTTTCCATT 120  
DB 1030 CGAGAGTGTGGCACCACATCGCTGAACACACACAGCATCTGTCTCCAGTTTCCATT 1089  
QY 121 GGGTCAGGATCATGTCGATTAGTTACTTCTGACCCAGCATCTGTCTATAGCCCAAG 180  
DB 1090 GGGTCAGGATCATGTCGATTAGTTACTTCTGACCCAGCATCTGTCTATAGCCCAAG 1149  
QY 181 AACAAATACCGTGTGATTGATTCAGTGGAGAAATAGAGCCCTTCCAATGTGAGCACA 240  
DB 1150 AACAAATACCGTGTGATTGATTCAGTGGAGAAATAGAGCCCTTCCAATGTGAGCACA 1209  
QY 241 GTCATCCACATCTGTACTCTCCGAGGAGCCAAAGAGGAGCGTGCAGTTCCAGTGG 300  
DB 1210 GTCATCCACATCTGTACTCTCCGAGGAGCCAAAGAGGAGCGTGCAGTTCCAGTGG 1269  
QY 301 AAACAGGACAGCCCTGCGAGTGGGTGAGGTGTATGAGGCCCTGCTGGGCCCTG 351  
DB 1270 AAACAGGACAGCCCTGCGAGTGGGTGAGGTGTATGAGGCCCTGCTGGGCCCTG 1320

Qy	1	GAGCAGTGTGGCACCACATCATGCGCATGCTGTCACTTCTGTGTAGCCGCTAGCGCCCT	60
Db	970	GAGCAGTGTGGCACCACATCATGCGCATGCTGTCACTTCTGTGTAGCCGCTAGCGCCCT	1029
Qy	61	CGAGAGCTGACCACACACATCGCTGGAACCAACAACAGCATCTGTCTCCAGTCTTTCCATT	120
Db	1030	CGAGAGCTGACCACACACATCGCTGGAACCAACAACAGCATCTGTCTCCAGTCTTTCCATT	1089
Qy	121	GGGTGAGGATCATGTGCGATTTTACTTCTGTACCCAGCATCATCTGTGTGCATACGCCAAG	180
Db	1090	GGGTGAGGATCATGTGCGATTTTACTTCTGTACCCAGCATCATCTGTGTGCATACGCCAAG	1149
Qy	181	AACAATACCGCTGATTTGGATTCAGCTGGAGAAAAATTAGAGCCCTTCCAATGTGAGCACA	240
Db	1150	AACAATACCGCTGATTTGGATTCAGCTGGAGAAAAATTAGAGCCCTTCCAATGTGAGCACA	1209
Qy	241	GTCTACACATCCTGTACCTCCCGAGGAGCAAGGGAGAGCGTGCAGTTCACGTGG	300
Db	1210	GTCTACACATCCTGTACCTCCCGAGGAGCAAGGGAGAGCGTGCAGTTCACGTGG	1269
Qy	301	AAACGAGCAGCGCTGCGAGTGGGTGATGAGGCGCTGCTGGGCCCTG	351
Db	1270	AAACGAGCAGCGCTGCGAGTGGGTGATGAGGCGCTGCTGGGCCCTG	1320
RESULT 4			
ABI99284			
ID	ABI99284 standard; cDNA; 11673 BP.		
XX			
AC	ABI99284;		
XX			
DT	07-MAR-2002 (first entry)		
XX			
DE	Mouse ischaemic condition related cDNA sequence SEQ ID NO:128.		
XX			
KW	Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;		
KW	vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.		
XX			
OS	Mus musculus.		
XX			
PN	W0200188188-A2.		
XX			
PD	22-NOV-2001.		
XX			
PF	18-MAY-2001; 2001WO-JP04192.		
XX			
PR	18-MAY-2000; 2000JP-0145977.		
XX			
PA	(UYNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.		
XX			
PI	Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;		
XX			
DR	WPI: 2002-034733/04.		
DR	P-PSDB; AB857065.		
XX			
PT	Examining the ischemic condition (e.g. occlusive ischemia) by measuring		
PT	expression levels of particular genes defined in the specification or		
PT	by determining the expression profile of a gene group comprising these		
PT	genes -		
XX			
PS	Claim 2; Page 362-385; 2690pp; English.		
XX			
CC	The present invention describes a method for examining ischaemic		
CC	conditions, comprising measuring the expression levels of particular		
CC	genes (I) in a test sample or determining the expression profile of a		
CC	gene group in the sample comprising genes selected from (I). The method		
CC	is useful for examining the ischaemic condition (e.g. compressive		
CC	ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring		
CC	expression levels of particular genes (ABI99202 to ABI99912, encoding		
CC	the protein sequences in AB857020 to AB857374) or by determining the		
CC	expression profile of a gene group comprising these genes. The		
CC	expression levels or expression profiles produced by these genes are		
CC	used as an indicator when screening for ischaemic condition-improving		

CC drugs or therapeutics for ischaemic diseases. AB199913 and AB199914  
 CC represent PCR primers for a mouse ischaemic condition related sequence,  
 CC which are used in the exemplification of the present invention.  
 XX  
 SQ Sequence 11673 BP; 2831 A; 2985 C; 2985 G; 2872 T; 0 other;  
 Query Match 100.0%; Score 351; DB 24; Length 11673;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-102;  
 Matches 351; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GAGCAGTGTGCACCATCATGCGCAATGCTGCTACCTTCTGTGAGCCGTACGGCCCT 60  
 DB 970 GAGCAGTGTGCACCATCATGCGCAATGCTGCTACCTTCTGTGAGCCGTACGGCCCT 1029  
 QY 61 CGAGAGTGTGCACCATCATGCGCAATGCTGCTACCTTCTGTGAGCCGTACGGCCCT 120  
 DB 1030 CGAGAGTGTGCACCATCATGCGCAATGCTGCTACCTTCTGTGAGCCGTACGGCCCT 1089  
 QY 121 GGGTCAGGATCATGCTGATTTAGTTACTCTGACCTTCTGTGAGCCGTACGGCCCT 180  
 DB 1090 GGGTCAGGATCATGCTGATTTAGTTACTCTGACCTTCTGTGAGCCGTACGGCCCT 1149  
 QY 181 AACATACCGCTGATTTAGTTACTCTGACCTTCTGTGAGCCGTACGGCCCT 240  
 DB 1150 AACATACCGCTGATTTAGTTACTCTGACCTTCTGTGAGCCGTACGGCCCT 1209  
 QY 241 GTCATCCACATCCTGTACCTCCCGAGAGAGCCGAGAGAGAGAGAGAGAGAGAGAG 300  
 DB 1210 GTCATCCACATCCTGTACCTCCCGAGAGAGCCGAGAGAGAGAGAGAGAGAGAGAG 1269  
 QY 301 AACAGGACAGCCGCGAGTGGGTGAGGTGATGAGCCCTGCTGGGCCCTG 351  
 DB 1270 AACAGGACAGCCGCGAGTGGGTGAGGTGATGAGCCCTGCTGGGCCCTG 1320  
 RESULT 5  
 ABN96939  
 ID ABN96939 standard; DNA; 11580 BP.  
 XX  
 AC  
 XX  
 XX  
 DT  
 XX  
 DE  
 XX  
 KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;  
 KW metastatic liver tumour; cytostatic; expression profile; disease state;  
 KW disease progression; drug toxicity; drug efficacy; drug metabolism.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200229103-A2.  
 XX  
 PD 11-APR-2002.  
 XX  
 PF 02-OCT-2001; 2001WO-US30589.  
 XX  
 PR 02-OCT-2000; 2000US-237054P.  
 XX  
 PA (GENE-) GENE LOGIC INC.  
 XX  
 PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;  
 XX WPI; 2002-426119/45.  
 XX  
 PT Diagnosing and detecting the progression of liver cancer,  
 PT hepatocellular carcinoma or metastatic liver tumor in a patient,  
 PT involves detecting the level of expression of two or more genes in a  
 PT liver tissue sample -  
 XX  
 XX Claim 1; SEQ ID NO 3437; 298pp; English.  
 PS  
 XX The invention relates to a novel method for diagnosing and detecting the

CC progression of liver cancer, hepatocellular carcinoma or metastatic liver  
 CC tumour in a patient, and differentiating metastatic liver cancer from  
 CC hepatocellular carcinoma in a patient, involving detecting the level of  
 CC expression of two or more genes represented in ABN93503-ABN97455 in a  
 CC tissue sample. The method of the invention has hepatotropic, and  
 CC cytostatic activity. The method is useful for diagnosing and detecting  
 CC the progression of liver cancer, hepatocellular carcinoma and metastatic  
 CC liver carcinoma in a patient. The method is useful for identifying  
 CC expression profiles which serve as useful diagnostic markers as well as  
 CC markers that can be used to monitor disease states, disease progression,  
 CC drug toxicity, drug efficacy and drug metabolism.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 11580 BP; 3014 A; 2696 C; 2753 G; 3116 T; 1 other;  
 Query Match 75.1%; Score 263.6; DB 24; Length 11580;  
 Best Local Similarity 84.6%; Pred. No. 2.8e-74;  
 Matches 296; Conservative 0; Mismatches 54; Indels 0; Gaps 0;  
 QY 1 GAGCAGTGTGCACCATCATGCGCAATGCTGCTACCTTCTGTGAGCCGTACGGCCCT 60  
 DB 860 GAACAGTGTGGCGCATATGCGCAATGCGTACCTTCTGTGAGCCGTACGGCCCT 919  
 QY 61 CGAGAGTGTGCACCATCATGCGCAATGCGTACCTTCTGTGAGCCGTACGGCCCT 120  
 DB 920 CGAAGACTGATTACCAACAGCCCTTAATACAAACAGCTTCTGTCCCAATTTCCATT 979  
 QY 121 GGGTCAGGATCATGCTGATTTAGTTACTCTGACCTTCTGTGAGCCGTACGGCCCT 180  
 DB 980 GGGTCAGGATCATGCTGATTTAGTTACTCTGACCTTCTGTGAGCCGTACGGCCCT 1039  
 QY 181 AACAAATACCGCTGATTTAGTTACTCTGAGCAATTTAGAGCCCTTCCAATGTGAGCACA 240  
 DB 1040 AATAACTCTCGGACTGATTTAGTTACTCTGAGCAATTTAGAGCCCTTCCAATGTGAGCACA 1099  
 QY 241 GTCATCCACATCCTGTACCTCCCGAGAGAGCCGAGAGAGAGAGAGAGAGAGAGAG 300  
 DB 1100 ATCATCCATATCTCTACCTTCTGTGAGGAGCCCAAGGGAGAGATGTCCAATTTCCATT 1159  
 QY 301 AACAGGACAGCCGCGAGTGGGTGAGGTGATGAGCCCTGCTGGGCCCT 350  
 DB 1160 AAGCAGGAAAATCTTCTGTAGGTGAAGTGTATGAGCCCTGCTGGGCCCT 1209  
 RESULT 6  
 ABA92604  
 ID ABA92604 standard; cDNA; 11580 BP.  
 XX  
 AC  
 ABA92604;  
 XX  
 DT 21-MAR-2002 (first entry)  
 XX  
 DE Human reelin encoding cDNA SEQ ID NO:4.  
 XX  
 KW Human; reelin; low density lipoprotein receptor; LDLR; neuroprotective;  
 KW extracellular glycoprotein; nootropic; antilipemic; Alzheimer's disease;  
 KW neurodegenerative disorder; neuronal regeneration; cognitive function;  
 KW lipid metabolism disease; memory; developmental disorder; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 176..10558  
 FT /\*tag= a  
 FT /product= "human reelin"  
 XX  
 PN US6323177-B1.  
 XX  
 PD 27-NOV-2001.  
 XX  
 PF 16-JUN-1999; 99US-0334220.



Db	980	GGGTGAGTTTCATGTCGCTTTAGTTATTTCAGACCCAGCATCATCGTGTATTATGCCAAG	1038
Qy	181	AAACAATACCGCTGATTGGATTACAGCTGGAGAAAATTAGAGCCCTTCCAATGTGAGCACA	240
Db	1040	AATAACTCTGGGACTGGATTACAGCTAGAGAAAATTAGAGCCCTTCCAATGTGAGCACA	1099
Qy	241	GTCAATCCACATCTGTACCTCCCGAGGAGCCAAAGGGGAGAGCGTGCAGTTCCAGTGG	300
Db	1100	ATCATCCATATCTCTACCTTCCCTGAGGAGGCCAAAGGGGAGATGTCCAATTTTCAGTGG	1159
Qy	301	AAACAGGACAGCTCGGAGTGGGTGAGGTGATGAGGCCTGCTGGGCGCT	350
Db	1160	AAGCAGGAAAATCTCGTGTAGGTGAAGTGTATGAAGCCTGCTGGGCGTT	1209
RESULT 8			
AAAD22779	standard; cDNA; 2025 BP.		
XX	AAAD22779;		
XX	26-FEB-2002 (first entry)		
XX	Xenopus laevis intact reelin cDNA.		
DE	XX		
XX	Frog; reelin; F-spondin domain; CR-50 epitope; gene therapy; agyria;		
KW	polymicrogyria; ectopic gray matter; ss.		
XX	XX		
OS	Xenopus laevis.		
XX	XX		
FH	Key	Location/Qualifiers	
FT	CDS	157..2025	
FT		/*tag= a	
FT		/product= "Xenopus intact reelin protein"	
FT		/transl_except= (pos:1606..1608, aa:Cys)	
FT		/transl_except= (pos:1639..1641, aa:Cys)	
FT		/note= "CDS does not include stop codon"	
FT		/partial	
PN	EP1149844-A2.		
XX	XX		
PD	31-OCT-2001.		
XX	XX		
PF	11-APR-2001; 2001EP-0303411.		
XX	XX		
PR	11-APR-2000; 2000JP-0109954.		
XX	XX		
PA	(RIKE ) RIKEN KK.		
XX	XX		
PI	Mikoshihira K, Tabata H, Nakajima K;		
XX	XX		
WP	WPI: 2002-019320/03.		
DR	P-PSDB; AAL13607.		
XX	XX		
PT	Novel truncated Reelin protein containing F-spondin domain and CR-50		
PT	recognition site of Reelin protein, but not having Reelin repeat site,		
PT	useful to treat diseases including agyria due to abnormal neuron		
PT	alignment -		
XX	XX		
XX	Example 1; Fig 1A; 47pp; English.		
PS	XX		
CC	The invention relates to a truncated Reelin protein comprising a		
CC	F-spondin domain and a CR-50 recognition site but no reelin protein		
CC	repeat site. Reelin is an essential molecule in developing a normal		
CC	laminated structure of cerebrum. The truncated reelin protein and its		
CC	DNA are useful for treating diseases including agyria, polymicrogyria		
CC	and ectopic gray matter due to abnormal neuronal alignment. Truncated		
CC	reelin protein DNA is useful in gene therapy. The present sequence is		
CC	a cDNA encoding Xenopus laevis intact reelin protein.		
SQ	Sequence 2025 BP; 541 A; 419 C; 460 G; 604 T; 1 other;		

Best Local Similarity		72.4%;	Pred. No. le-52;
Matches	254;	Conservative	0; Mismatches 97; Indels 0; Gaps
Qy	1	GAGCAGTGTGGCACCATCATGCGATGCGCAATGCTGTCACTTCTGTGAGCCGTACGGCCCT	60
Db	847	GAGCAGTGTGGTGCATTAATGCAATTAATGCAATGCTGGGGCAGTCACFTTTGTGATCCCGTATGGACCA	906
Qy	61	CGAGAGCTGACCAACACACATGCCCTGAACACAAACACAGACATCTGTCTCCAGTTTCCATT	120
Db	907	AGACAATGTGATACTGTTCAATGACACACACTACGGCATCTGTTTTGCAGTTTCTATT	966
Qy	121	GGGTGAGGATCATGTGCAATTTAGTTACTCTGACCCAGCAGCATCTGTCTCATACGCCAAG	180
Db	967	GGGTGAGGATCGTGCAAGTTTCAGCTATTTCAGACCCCTGGAATTTGCTGTCTATACACAAG	1026
Qy	181	AACAATACCCCTGATTTGGATTTCAGCTGGAGAAAAATTAGAGCCCTTCCAATGTGAGCACA	240
Db	1027	AATAATTCATCTAGTTGGATGGCATTTGGAGAAATTAGTGTCTTCCAATGTTAGCACC	1086
Qy	241	GTATCCACATCTGTACCTCCCGAGGAGGCCAAAGGGGAGAGCGTGCAGTTCCAGTGG	300
Db	1087	ATCAATTCACATTAATTTACCTACCTCTGAAGCTAAAGGAGAAATGTGAATTCGTTGG	1146
Qy	301	AAACAGGACACCGCTGCCAGTGGGTGAGTGTATGAGCCCTGCTGGGCCCTG	351
Db	1147	AGGCAGGAGAACATGCAGGCAGGTGATGTGTATGAAGCCTGCTGGGCAC	1197
RESULT 9			
AAD22753			
ID	AAD22753 standard; cDNA; 2274 BP.		
AC	AAD22753;		
DT	26-FEB-2002 (first entry)		
DE	Xenopus laevis truncated reelin cDNA.		
Kw	Frog; reelin; F-spondin domain; CR-50 epitope; gene therapy; agyria;		
KW	polymicrogyria; ectopic gray matter; ss.		
OS	Xenopus laevis.		
Key	Location/Qualifiers		
FT	157..1455		
FT	/tag= a		
FT	/product= "X. laevis truncated reelin protein"		
FT	157..234		
FT	/tag= b		
FT	235..1452		
FT	/tag= c		
FT	/product= "Mature truncated reelin protein"		
FT	241..726		
FT	/tag= d		
FT	/note= "Encodes F-spondin domain"		
FT	847..1197		
FT	/tag= e		
FT	/note= "Encodes CR-50 epitope region"		
FT	2254..2259		
FT	/tag= f		
FN	EPI149844-A2.		
PD	31-OCT-2001.		
XX	11-APR-2001; 2001EP-0303411.		
XX	11-APR-2000; 2000JP-0109954.		
XX	(RIKE ) RIKEN KK.		
XX	Mikoshiba K, Tabata H, Nakajima K;		
XX			

DR WPI: 2002-019320/03.  
XX P-PSDB; AAE13605.  
PT Novel truncated Reelin protein containing F-spondin domain and CR-50  
PT recognition site of Reelin protein, but not having Reelin repeat site,  
PT useful to treat diseases including agyria due to abnormal neuron  
PT alignment -  
XX  
XX  
PS Claim 8; Page 11-16; 47pp; English.  
XX  
XX The invention relates to a truncated Reelin protein comprising a  
CC F-spondin domain and a CR-50 recognition site but no reelin protein  
CC repeat site. Reelin is an essential molecule in developing a normal  
CC laminated structure of cerebrum. The truncated reelin protein and its  
CC DNA are useful for treating diseases including agyria, polymicrogyria,  
CC and ectopic gray matter due to abnormal neuronal alignment. Truncated  
CC reelin protein DNA is useful in gene therapy. The present sequence is  
CC a cDNA encoding Xenopus laevis truncated reelin protein.  
XX  
XX Sequence 2274 BP; 660 A; 450 C; 469 G; 694 T; 1 other;  
SQ  
Query Match 55.8%; Score 195.8; DB 24; Length 2274;  
Best Local Similarity 72.4%; Pred. No. 1e-52;  
Matches 254; Conservative 0; Mismatches 97; Indels 0; Gaps 0;  
QY 1 GAGCAGTGTGGCACCACATCATGCGCAATGCTGTCACTTCTGTGAGCCGTACGGCCCT 60  
DB 847 GAGCAGTGTGTGCAATATATGCGATGGTGGGCGACGTCACTTTTGTGATCGGTATGGACCA 906  
QY 61 CGAGAGCTGACCCACATCGCTGAACACAAACACAGCATCTGTCCCTCCAGTTTTCATT 120  
DB 907 AGAATTTGATACTGTTCAATGAACACAACTACGCACTGTTCGACGTTTCTATT 966  
QY 121 GGTGAGATCATGTCGATTAGTTACTGTGACCCAGCATCATGTGTATGATGCGCAAG 180  
DB 967 GGTGAGATCATGTCGATTAGTTACTGTGACCCAGCATCATGTGTATGATGCGCAAG 1026  
QY 181 AACAAATCCGCTGATTGATTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAG 240  
DB 1027 AATAATTCATCTAGTTGGATGCCATTGGAGAGAAATAGTCTCTCCATTTAGTACCC 1086  
QY 241 GTATCCACATCTCTGTACCTCCCGAGGAGCCAAAGGAGAGCGTGCAGTTCAGTGG 300  
DB 1087 ATCATTCACATTTATTTACCTACCTCTGAGCTAAGAGAGAAATGTGAATTCCTGTTG 1146  
QY 301 AAACAGGACAGCCCTGCGAGTGGGTGAGGTGTATGAGGCGCTGCTGGGCCCTG 351  
DB 1147 AGGCAGGAGAACATGTCAGGCGAGTGTGTGTATGAAGCCTGCTGGGCACTG 1197  
RESULT 10  
AAK07720/c  
ID AAK07720 standard; DNA; 413 BP.  
XX  
AC AAK07720;  
XX  
XX  
DT 05-NOV-2001 (first entry)  
XX  
XX Human brain expressed single exon probe SEQ ID NO: 7711.  
DE  
XX  
KW Human; brain expressed exon; gene expression analysis; probe;  
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
KW epilepsy; cancer; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO200157275-A2.  
PN  
XX  
XX 09-AUG-2001.  
PD  
XX  
XX 30-JAN-2001; 2001WO-US00667.  
PF  
XX  
XX 04-FEB-2000; 2000US-0180312.  
PR  
DR WPI: 2002-019320/03.  
XX P-PSDB; AAE13605.  
PT Novel truncated Reelin protein containing F-spondin domain and CR-50  
PT recognition site of Reelin protein, but not having Reelin repeat site,  
PT useful to treat diseases including agyria due to abnormal neuron  
PT alignment -  
XX  
XX  
PS Claim 8; Page 11-16; 47pp; English.  
XX  
XX The invention relates to a truncated Reelin protein comprising a  
CC F-spondin domain and a CR-50 recognition site but no reelin protein  
CC repeat site. Reelin is an essential molecule in developing a normal  
CC laminated structure of cerebrum. The truncated reelin protein and its  
CC DNA are useful for treating diseases including agyria, polymicrogyria,  
CC and ectopic gray matter due to abnormal neuronal alignment. Truncated  
CC reelin protein DNA is useful in gene therapy. The present sequence is  
CC a cDNA encoding Xenopus laevis truncated reelin protein.  
XX  
XX Sequence 2274 BP; 660 A; 450 C; 469 G; 694 T; 1 other;  
SQ  
Query Match 55.8%; Score 195.8; DB 24; Length 2274;  
Best Local Similarity 72.4%; Pred. No. 1e-52;  
Matches 254; Conservative 0; Mismatches 97; Indels 0; Gaps 0;  
QY 1 GAGCAGTGTGGCACCACATCATGCGCAATGCTGTCACTTCTGTGAGCCGTACGGCCCT 60  
DB 847 GAGCAGTGTGTGCAATATATGCGATGGTGGGCGACGTCACTTTTGTGATCGGTATGGACCA 906  
QY 61 CGAGAGCTGACCCACATCGCTGAACACAAACACAGCATCTGTCCCTCCAGTTTTCATT 120  
DB 907 AGAATTTGATACTGTTCAATGAACACAACTACGCACTGTTCGACGTTTCTATT 966  
QY 121 GGTGAGATCATGTCGATTAGTTACTGTGACCCAGCATCATGTGTATGATGCGCAAG 180  
DB 967 GGTGAGATCATGTCGATTAGTTACTGTGACCCAGCATCATGTGTATGATGCGCAAG 1026  
QY 181 AACAAATCCGCTGATTGATTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAG 240  
DB 1027 AATAATTCATCTAGTTGGATGCCATTGGAGAGAAATAGTCTCTCCATTTAGTACCC 1086  
QY 241 GTATCCACATCTCTGTACCTCCCGAGGAGCCAAAGGAGAGCGTGCAGTTCAGTGG 300  
DB 1087 ATCATTCACATTTATTTACCTACCTCTGAGCTAAGAGAGAAATGTGAATTCCTGTTG 1146  
QY 301 AAACAGGACAGCCCTGCGAGTGGGTGAGGTGTATGAGGCGCTGCTGGGCCCTG 351  
DB 1147 AGGCAGGAGAACATGTCAGGCGAGTGTGTGTATGAAGCCTGCTGGGCACTG 1197  
RESULT 10  
AAK07720/c  
ID AAK07720 standard; DNA; 413 BP.  
XX  
AC AAK07720;  
XX  
XX  
DT 05-NOV-2001 (first entry)  
XX  
XX Human brain expressed single exon probe SEQ ID NO: 7711.  
DE  
XX  
KW Human; brain expressed exon; gene expression analysis; probe;  
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
KW epilepsy; cancer; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO200157275-A2.  
PN  
XX  
XX 09-AUG-2001.  
PD  
XX  
XX 30-JAN-2001; 2001WO-US00667.  
PF  
XX  
XX 04-FEB-2000; 2000US-0180312.  
PR

26-MAY-2000; 2000US-0207456.  
30-JUN-2000; 2000US-0608408.  
03-AUG-2000; 2000US-0632366.  
21-SEP-2000; 2000US-0234687.  
27-SEP-2000; 2000US-0236359.  
04-OCT-2000; 2000GB-0024263.  
(MOLE-) MOLECULAR DYNAMICS INC.  
Penn SG, Hanzel DK, Chen W, Rank DR;  
WPI: 2001-483446/52.  
Single exon nucleic acid probes for analyzing gene expression in human brains -  
Example 4; SEQ ID NO: 7711; 650pp + Sequence Listing; English.  
The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human brain. They can be used to measure gene expression in brain cell samples, which may enable the diagnosis and improved treatment of nervous system diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, epilepsy and cancers. The present sequence is one of the probes of the invention.  
Sequence 413 BP; 134 A; 86 C; 67 G; 126 T; 0 other;  
Query Match 15.3%; Score 53.8; DB 22; Length 413;  
Best Local Similarity 79.0%; Pred. No. 3.4e-07;  
Matches 64; Conservative 0; Mismatches 17; Indels 0; Gaps 0;  
QY 1 GAGCAGTGTGGCACCACATCATGCGCAATGCTGTCACTTCTGTGAGCCGTACGGCCCT 60  
DB 323 GAACAGTGTGGCGGATATGCGATGGCAATGCCGTCACTTCTGTGAACCATATGGCCCA 264  
QY 61 CGAGAGCTGACCCACATCATGCGCAATGCTGTCACTTCTGTGAGCCGTACGGCCCT 81  
DB 263 CGAGAACTGGTAAGTATGTGC 243  
RESULT 11  
AAK33553/c  
ID AAK33553 standard; DNA; 413 BP.  
XX  
AC AAK33553;  
XX  
XX 06-NOV-2001 (first entry)  
XX  
XX Human bone marrow expressed single exon probe SEQ ID NO: 8110.  
DE  
XX  
KW Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO200157276-A2.  
PN  
XX  
XX 09-AUG-2001.  
PD  
XX  
XX 30-JAN-2001; 2001WO-US00668.  
PF  
XX  
XX 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
(MOLE-) MOLECULAR DYNAMICS INC.  
Penn SG, Hanzel DK, Chen W, Rank DR;  
WPI: 2001-483446/52.  
Single exon nucleic acid probes for analyzing gene expression in human brains -  
Example 4; SEQ ID NO: 7711; 650pp + Sequence Listing; English.  
The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human brain. They can be used to measure gene expression in brain cell samples, which may enable the diagnosis and improved treatment of nervous system diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, epilepsy and cancers. The present sequence is one of the probes of the invention.  
Sequence 413 BP; 134 A; 86 C; 67 G; 126 T; 0 other;  
Query Match 15.3%; Score 53.8; DB 22; Length 413;  
Best Local Similarity 79.0%; Pred. No. 3.4e-07;  
Matches 64; Conservative 0; Mismatches 17; Indels 0; Gaps 0;  
QY 1 GAGCAGTGTGGCACCACATCATGCGCAATGCTGTCACTTCTGTGAGCCGTACGGCCCT 60  
DB 323 GAACAGTGTGGCGGATATGCGATGGCAATGCCGTCACTTCTGTGAACCATATGGCCCA 264  
QY 61 CGAGAGCTGACCCACATCATGCGCAATGCTGTCACTTCTGTGAGCCGTACGGCCCT 81  
DB 263 CGAGAACTGGTAAGTATGTGC 243

```

XX WPI; 2001-488900/53.
DR
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
PT
XX Example 4; SEQ ID NO: 8110; 658pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
SQ Sequence 413 BP; 134 A; 86 G; 67 C; 126 T; 0 other;

Query Match          15.3%; Score 53.8; DB 22; Length 413;
Best Local Similarity 79.0%; Pred. No. 3.4e-07;
Matches 64; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY      1   GAGCAGTGTGGCACCATCATGCATGCGCAATGCTGTACACTTCCTGTGAGCCGTACGGCCCT 60
Db      323 GAACAGTGTGGCGGATTATGATGGTAATGCCGTACCTTCTGTGAACCATATGCCCCA 264
        ||||| ||| | | ||||| ||||| ||||| ||||| ||||| ||| || |||||
QY      61   CGAGAGCTGACCACCATGC 81
Db      263 CGAGAAGTGTAAGTAGTGTC 243
        ||||| ||| | | |||

RESULT 12
ABS08422/c
ID ABS08422 standard; DNA; 413 BP.
XX
AC ABS08422;
XX
DT DT
XX
XX 19-AUG-2002 (first entry)
DE
XX Human genome-derived single exon probe from lung SEQ ID No 8413.
KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
KW chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagenen syndrome;
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW primary ciliary dyskinesia; pulmonary hypertension;
KW hyaline membrane disease.
XX
OS Homo sapiens.
XX
PN WO200186003-A2.
XX
PD 15-NOV-2001.
XX
PF 30-JAN-2001; 2001WO-US000665.
XX
PR 04-FEB-2000; 2000US-180312P.
PR 26-MAY-2000; 2000US-207456P.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-234687P.
PR 27-SEP-2000; 2000US-236359P.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2002-114183/15.
XX
PT Spatially-addressable set of single exon nucleic acid probes, used to

```

measure gene expression in human lung samples -

Claim 1; SEQ ID NO 8413; 634pp; English.

The invention relates to a spatially-addressable set of single exon nucleic acid probes for measuring gene expression in a sample derived of from human lung comprising single exon nucleic acid probes having one of 12614 nucleic acid sequences mentioned in the specification, or their complements or the 12387 open reading frames derived from the 12614 probes. Also included are a microarray comprising the novel set of probes; the novel set of probes which hybridise at high stringency to a nucleic acid expressed in the human lung; measuring gene expression to a sample derived from human lung, comprising (a) contacting the array with a collection of detectably labeled nucleic acids derived from human lung mRNA, and (b) measuring the label detectably bound to each probe of the array; identifying exons in a eukaryotic genome, comprising (a) algorithmically predicting at least one exon from genomic sequences of the eukaryote; and (b) detecting specific hybridisation of detectably labeled nucleic acids from eukaryote lung mRNA, to a single exon probe, having a fragment identical to the predicted exon, the probe is included in the above mentioned microarray; assigning exons to a single gene, comprising (a) identifying exons from genomic sequence by the method above and (b) measuring the expression of each of the exons in several tissues and/or cell types using hybridisation to a single exon microarrays having a probe with the exon, where a common pattern of expression of the exons in the tissues and/or cell types indicates that the exons should be assigned to a single gene; a peptide comprising one of 12011 sequences, mentioned in the specification, or encoded by the probes/open reading frames (ORF). The probes are used for gene expression analysis, and for identifying exons in a gene, particularly using human lung derived mRNA and for the study of lung diseases such as asthma, lung cancer, chronic obstructive pulmonary disease (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary haemorrhoidosis, pulmonary histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension and hyaline membrane disease. The present sequence is a single exon probe of the invention.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from wipo at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).

Sequence 413 BP; 134 A; 86 C; 67 G; 126 T; 0 other;

Query Match 15.3%; Score 53.8; DB 24; Length 413;  
Best Local Similarity 79.0%; Pred. No. 3.4e-07;  
Matches 64; Conservative 0; Mismatches 17; Indels 0; Gaps 0

QY 1 GAGCAGTGTGGCACCACATCATGCATGTCACCTCTGTGAGCCGTACGGCCT 60  
||| ||||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 323 GAACAGTGTGGCGCATTTATGCATGCATGCATGCCTCTGTGAAACCATATGCGCCA 264

QY 61 CGAGACGTGACCACCATGCC 81  
||| |||| ||| ||| ||| |||

Db 263 CGAGAACTGGTAAGTATGTC 243

RESULT 13  
ID AAK20384/C  
XX AAK20384 standard; DNA; 87 BP.  
XX AAK20384;  
XX  
XX  
DT 05-NOV-2001 (first entry)  
XX  
DE Human brain expressed single exon probe SEQ ID NO: 20375.  
XX  
XX Human; brain expressed exon; gene expression analysis; probe;  
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
KW epilepsy; cancer; ss.







GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: November 6, 2002, 13:54:36 ; Search time 2213 Seconds  
(without alignments)  
2568.737 Million cell updates/sec

Title: US-09-897-438B-1  
Perfect score: 351  
Sequence: 1 gagcagtggccaccatcat.....atgagccctgctggccctg 351

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST.\*  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estnu.\*  
4: em\_estnu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_estro.\*  
9: gb\_esti.\*  
10: gb\_est2.\*  
11: gb\_hic.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estom.\*  
17: gb\_gss.\*  
18: em\_gss\_hum.\*  
19: em\_gss\_inv.\*  
20: em\_gss\_pln.\*  
21: em\_gss\_vrt.\*  
22: em\_gss\_fun.\*  
23: em\_gss\_mam.\*  
24: em\_gss\_mus.\*  
25: em\_gss\_other.\*  
26: em\_gss\_pro.\*  
27: em\_gss\_rod.\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	203.8	58.1	315	12	BE938667 QV0-TN008
C 2	145	41.3	600	12	BG803882 0243-03 M
C 3	94	26.8	635	10	BB248113 BB248113
C 4	64	18.2	568	12	BF387628 UT-R-CAL-
C 5	39.6	11.3	943	14	BQ647080 AGENCOURT
C 6	39.2	11.2	514	14	BQ955513 QGG10D11.

7	38.6	11.0	1022	17	CNS012TZ	AL174752 Tetraodon
8	36	10.3	547	13	CNS03750	BI803750 H18A05 E
C 9	35.6	10.1	993	17	BI804W64	AL310333 Tetraodon
C 10	35.4	10.1	330	10	AW922767	AW922767 DGI_46_E0
C 11	34.8	9.9	374	17	AZ777615	AZ777615 2M0012G10
C 12	34.8	9.9	582	10	AV522139	AV522139 AV522139
C 13	34.8	9.9	668	13	BI573885	BI573885 RH09360.3
C 14	34.6	9.9	528	14	BQ038901	BQ038901 pgnic.pk0
C 15	34.6	9.9	917	12	BF135143	BF135143 601780186
C 16	34.2	9.7	507	10	AW920326	AW920326 EST351630
C 17	34	9.7	505	14	BM956552	BM956552 LM24HW006
C 18	34	9.7	547	12	BE08801	BE08801 214079 MA
C 19	33.8	9.6	872	17	CNS03DM3	AL239520 Tetraodon
C 20	33.6	9.6	435	17	AZ558748	AZ558748 RPCI-23-1
C 21	33.4	9.5	563	12	BG420388	BG420388 602452417
C 22	33.4	9.5	1077	17	CNS04XQV	AL312016 Tetraodon
C 23	33.2	9.5	476	12	BF499004	BF499004 AT11790.5
C 24	33.2	9.5	485	17	AZ822150	AZ822150 2M0095C09
C 25	33.2	9.5	691	12	BG600888	BG600888 EST505795
C 26	33.2	9.5	715	13	BM622337	BM622337 17006874
C 27	33.2	9.5	725	12	BF492454	BF492454 AT25588.5
C 28	33	9.4	332	10	AW401228	AW401228 Lamdigest
C 29	33	9.4	668	17	AQ475926	AQ475926 CITBI-E1-
C 30	32.8	9.3	412	12	BF288465	BF288465 EST453056
C 31	32.6	9.3	589	14	BM715558	BM715558 UI-E-EJO-
C 32	32.6	9.3	874	9	AL554368	AL554368 AL554368
C 33	32.6	9.3	938	17	CNS03BLG	AL236653 Tetraodon
C 34	32.4	9.2	419	17	AQ237938	AQ237938 RPCI11-65
C 35	32.4	9.2	431	12	BF603741	BF603741 269316 MA
C 36	32.4	9.2	465	9	AA835142	AA835142 ak63906.s
C 37	32.4	9.2	701	17	AQ316104	AQ316104 RPCI11-10
C 38	32.4	9.2	851	17	BH129632	BH129632 G-5J5 Mai
C 39	32.4	9.2	1038	17	CNS05S4W	AL351401 Tetraodon
C 40	32.2	9.2	412	9	AA617544	AA617544 vj76a10.r
C 41	32.2	9.2	483	12	BF441217	BF441217 257407 MA
C 42	32.2	9.2	719	10	BE614332	BE614332 601504394
C 43	32	9.1	455	10	AW566596	AW566596 fK02404.Y
C 44	32	9.1	550	13	BI865114	BI865114 fC97e01.Y
C 45	32	9.1	559	13	BI318806	BI318806 fQ65c09.Y

## ALIGNMENTS

RESULT 1  
BE938667/C  
LOCUS BE938667 315 bp mRNA linear EST 02-OCT-2000  
DEFINITION QV0-TN0084-180800-342-a08 TN0084 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BE938667  
VERSION BE938667.1 GI:10466224  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 315)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare M.J., Soares,F., Brentani,R., Reis,L.F., de Souza,S.J. and Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

TITLE  
JOURNAL  
MEDLINE  
COMMENT  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001







```

SOURCE      Tetraodon nigroviridis.
ORGANISM    Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE   1 (bases 1 to 1022)
AUTHORS    Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
TITLE      Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
JOURNAL    Unpublished
AUTHORS    2 (bases 1 to 1022)
            Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
TITLE      Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
JOURNAL    Unpublished
AUTHORS    3 (bases 1 to 1022)
            Direct Submission
            Submitted (12-APR-2000)
COMMENT    This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES   1..1022
            source
            /organism="Tetraodon nigroviridis"
            /db_xref="taxon:99883"
            /clone_lib="G"
            /note="Genoscope sequence ID : COAG221AG08SP1-end :
            PUC-Orig"
BASE COUNT 246 a 227 c 241 g 307 t 1 others
ORIGIN
Query Match 11.0%; Score 38.6; DB 17; Length 1022;
Best Local Similarity 72.5%; Pred. No. 1.3;
Matches 50; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
QY 1 GAGCAGTGTGCACCATCATCGATGGCAATGCTGTACCTTCTGTGAGCCGTAGGCCCT 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 156 GAGCAGTGTGGGGTCCATCATCGATGGACGCGGTGACCTTCTGTGAACCCCTTGGAG 215
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 CGAGAGCTG 69
    ||||| |||||
Db 216 CGAGAGCTG 224

RESULT 8
BI803750
LOCUS      BI803750 547 bp mRNA linear EST 01-NOV-2001
DEFINITION H118A05 Endosperm library from Oryza sativa (10 days after anthesis
            ) Oryza sativa cDNA clone H118A05, mRNA sequence.
ACCESSION  BI803750
VERSION     BI803750.1 GI:16576454
KEYWORDS    EST.
SOURCE      Oryza sativa.
ORGANISM    Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1 (bases 1 to 547)
AUTHORS    Dong,H.T., Li,D.B., Zhuang,X.F., Dai,C.G., Sun,L.X., Pei,Y.X., Wu
            ,H.F., Jiang,Y.X., Yu,F.C., Gao,Q.K. and Lou,Y.C.
TITLE      A Gene Expression Screen in Oryza sativa
JOURNAL    Unpublished (2001)
COMMENT    Contact: Dong HT
            Laboratory of Functional Genetics
            Bio-technology Institute of Zhejiang University

```

```

FEATURES   1..547
            source
            /organism="Oryza sativa"
            /db_xref="taxon:4530"
            /clone="H118A05"
            /clone_lib="Endosperm library from Oryza sativa (10 days
            after anthesis)"
            /tissue_type="Endosperm"
            /dev_stage="10 days after anthesis"
            /note="vector: pSport2"
BASE COUNT 119 a 144 c 148 g 136 t
ORIGIN
Query Match 10.3%; Score 36; DB 13; Length 547;
Best Local Similarity 51.2%; Pred. No. 5.9;
Matches 84; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 153 CCCGAGCATCACTGTGTATACATACGCGCAAGACAATACCGCTGATTGGATTTCAGCTGGAGAA 212
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 164 CCACTCCCTCCCATCTCTCTCTCCAGATCTGGCAGAGGGGGAAGTAGGTGGCGGC 223
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 213 AATTAGACCCCTTCCCAATGTGAGCAGATCATCCATCCTGTACCTCCCGGAGGAAGC 272
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 224 GACAAAGAGAGTGCGAAGGACGCGCTTCTCTCCCTTCTCTCTCAGATCCAAC 283
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 273 CAAGGGGAGAGCGTGCAGTTCAGTTCAGTGGAAACAGACAGCCCTGC 316
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 284 CGAGGGGGGAGCGCGGTGGCAACTGCACCGTAGAGGCGAGC 327

RESULT 9
CNS04WGA/C
LOCUS      CNS04WGA 993 bp DNA linear GSS 26-JUL-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
            012G24 of library A from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION  AL310333.1 GI:9543201
VERSION     AL310333
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
ORGANISM    Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE   1 (bases 1 to 993)
AUTHORS    Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
TITLE      Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
JOURNAL    Nat. Genet. 25 (2), 235-238 (2000)
MEDLINE    20296633
PUBMED     10835645
REFERENCE   2 (bases 1 to 993)
AUTHORS    Crollius,H.R., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C.,
            Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W.,
            Bernot,A. and Weissenbach,J.
TITLE      Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
JOURNAL    Genome Res. 10 (7), 939-949 (2000)
MEDLINE    20359837
PUBMED     10896143
REFERENCE   3 (bases 1 to 993)
AUTHORS    Direct Submission
TITLE      Submitted (12-APR-2000)
COMMENT    This sequence is a single read and was generated as part of a large

```

Kaixuan Road 268#, Hangzhou, Zhejiang, P.R.China  
 Tel: 0086-571-86892051  
 Fax: 0086-571-86961525  
 Email: htdong@zjuem.zju.edu.cn  
 Seq primer: M13 forward primer.

scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

## FEATURES

## source

1..993

/organism="Tetraodon nigroviridis"

/db\_xref="taxon:99883"

/clone="012G24"

/clone\_lib="A"

/note="Genoscope sequence ID : COAA012BD12C1-end : T7"

214 a 272 c 229 g 238 t 40 others

## BASE COUNT

## ORIGIN

## Query Match

Best Local Similarity 10.1%; Score 35.6; DB 17; Length 993;

Matches 86; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

Qy 60 TCGAGAGTGACACCATCCCTGACACACAAACAGCAGCTGTCTCCAGTTTCAT 119

Db 773 TCCTGGCGTGAGCTCATCCGNNATGCGTCAACAAAGCAGAGCNCCTTCAAGGACNCCT 714

Qy 120 TGGGTGAGGATCATGTCTAGTTAGTTACTCTGACCCAGCAGCATCACTGTGTCATACGCCAA 179

Db 713 TTCTAACACGCTCTCTATTNATGGCTACGATCAGCGATTNANGCCGTTCCACACG 654

Qy 180 GAACAATACCGCTGATTGATTCAGCTGGAGAAAATTAGAGCCCTTCCCAATGTGAGCA 238

Db 653 GAAGGACAGCGCCGCTCGACANACAATGAGNATATTTCCCCGCTCACAATGCGGCA 595

## RESULT 10

## AW922767/c

## LOCUS

DEFINITION AW922767 330 bp mRNA linear EST 19-JUL-2000

DG1\_46.E09.bl\_A002 Dark Grown 1 (DG1) Sorghum bicolor cDNA, mRNA

sequence.

## ACCESSION

AW922767

## VERSION

AW922767.1

## KEYWORDS

EST.

## SOURCE

Sorghum bicolor

## ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

clade; Panicoideae; Andropogoneae; Sorghum.

1 (bases 1 to 330)

Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt

,L.H.

An EST database from Sorghum: dark-grown seedlings

Unpublished (2000)

Contact: Cordonnier-Pratt MM

Laboratory for Genomics and Bioinformatics

The University of Georgia, Department of Plant Biology

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 583 0210

Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions

below Phred quality 16. The threshold for highest quality sequence

is 20.

Seq primer: JEN REV

High quality sequence stop: 326

POLYA-No.

Location/Qualifiers

1..330

/organism="Sorghum bicolor"

/db\_xref="taxon:4558"

/clone\_lib="Dark Grown 1 (DG1)"

/note="Organ: 5-day-old dark-grown seedlings; Vector:

Lambda Zap; Site\_1: XhoI; Site\_2: EcoRI; The library was

made from poly-A RNA in the cloning vector lambda Zap II.

Clones to be sequenced were prepared by mass excision."

53 a 95 c 108 g 74 t

BASE COUNT

ORIGIN

## Query Match

Best Local Similarity 10.1%; Score 35.4; DB 10; Length 330;

Matches 66; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy 146 ACTCTGACCCAGCATCTGTGTCTATACGCCAAGACATACCGCTGATTGGATTTCAGC 205

Db 216 ACTTTGATACACACCCGCGGAAACAGCATGCGCTGCGGTGGCGGAACACCA 157

Qy 206 TGGAGAAATAGAGCCCTTCCCAATGTGACACAGCATCCACATCCTCTPACCTCC 262

Db 156 TTCTGAAATCCCGACGACGCGATGAGATCGCAGTCTCCACATGTTGGAGTGC 100

## RESULT 11

## AZ777615

## LOCUS

DEFINITION 374 bp DNA linear GSS 16-FEB-2001

clone UUGC2M0012G10 F, DNA sequence.

AZ777615

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0012 row: G column: 10

Seq primer: CGTTGTAACACGACGCGCCAGT

Class: plasmid ends

High quality sequence stop: 374.

Location/Qualifiers

1..374

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC2M0012G10"

/clone\_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(<http://www.jax.org/resources/documents/dnares/>). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of pWD42 (g114732114|gb|AF129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to

adapted vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells





```
DEFINITION   pgnlc.pk010.m22 normalized chicken lymphoid cdna library Gallus
              gallus cDNA clone pgnlc.pk010.m22 5' similar to emb|CAC84904.1
              (AJ317960) transposition factor [Gallus gallus], mRNA sequence.
ACCESSION    BQ038901
VERSION      BQ038901.2  GI:20383663
KEYWORDS     EST.
SOURCE       chicken.
ORGANISM     Gallus gallus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
              Phasianinae; Gallus.
REFERENCE    1 (bases 1 to 528)
AUTHORS      Morgan,R.W. and Burnside,J.
TITLE        Chicken lymphoid ESTs
JOURNAL      Unpublished (2001)
COMMENT      On Mar 27, 2002 this sequence version replaced gi:19772441.
              Contact: Joan Burnside
              Molecular Endocrinology
              University of Delaware
              40 Townsend Hall, Newark, DE 19717, USA
              Tel: 302-831-1345
              Fax: 302-831-3411
              Email: joan@udel.edu, www.chickest.udel.edu.
FEATURES     Location/Qualifiers
              source
                1..528
                /organism="Gallus gallus"
                /db_xref="taxon:9031"
                /clone="pgnlc.pk010.m22"
                /clone.lib="normalized chicken lymphoid cdna library"
                /sex="Male and Female"
                /tissue_type="thymus, bursa, spleen, PBL, bone marrow"
                /lab_host="E.coli EMDH10B"
                /note="Vector: pCMVSPORT 6"
BASE COUNT   110 a 202 c 124 g 79 t 13 others
ORIGIN
Query Match      9.9%; Score 34.6; DB 14; Length 528;
Best Local Similarity 44.2%; Pred. No. 15;
Matches 118; Conservative 0; Mismatches 149; Indels 0; Gaps 0;
QY  48  GCCGTACGGGCGCTCGAGAGCTGACACACATGCTGAACACACACACATCTGTCT 107
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  244  GCTAGCGGCTCCGGCTCCCGGACACAGCTCCACGCCACACCTCAGTGTGCGCAT 303
QY  108  CCAGTTTTCATTGGTTCAGAGTCATGTCGATTTAGTTACTCTGACCCAGCATCACTGT 167
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  304  CAGCTCGTCCAACATGCCACATCCACCGAGTTCGGGCTCTTTTTCACCAAGGCCAA 363
QY  168  GTCATAGCCCAAGAACAAATACCGTGTATTCGATTCAGCTGGAGAAATAGAGCCCTTC 227
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  364  CAGCTATTCCACCACCACCACCACCATGATGGGCATCATGAATTTTCCAGCAGTTC 423
QY  228  CAATGTGAGCAGTCATCCACATCTCTGTACCTCCCGGAGAGCCAAAGGGGAGACCGT 287
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  424  CACAGTGGGACAAACCTCCAAAGTGCAGACGCCCCAGNNNGCCCCAGCAGGAGTCCCT 483
QY  288  GCAGTTCCAGTGGAAACAGGACAGCCT 314
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  484  GCAGAGCCAGGTCNNNNNGNTGGCT 510

RESULT 15
BF135143/c
LOCUS
DEFINITION   BF135143 917 bp mRNA linear EST 24-OCT-2000
              601780186F1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:4008333 5',
              mRNA sequence.
ACCESSION    BF135143
VERSION      BF135143.1  GI:10974183
KEYWORDS     EST.
SOURCE       house mouse.
ORGANISM     Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE    1 (bases 1 to 917)
              NIH-MGC http://mgc.nci.nih.gov/.
              National Institutes of Health, Mammalian Gene Collection (MGC)
              Unpublished (1999)
              Contact: Robert Strausberg, Ph.D.
              Email: cgapbs@email.nih.gov
              Tissue procurement: Gilbert Smith, Ph.D.
              cDNA Library Preparation: Life Technologies, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: NCI-CGAP clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LIAW9242 row: k column: 22
              High quality sequence stop: 708.
FEATURES     Location/Qualifiers
              source
                1..917
                /organism="Mus musculus"
                /strain="CZECH II"
                /db_xref="taxon:10090"
                /clone.lib="IMAGE:4008333"
                /clone.lib="NCI_CGAP_Lu30"
                /tissue_type="tumor, metastatic to mammary"
                /lab_host="DH10B"
                /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
                Site_2: SalI; transgenic model WNT-1, expression driven by
                MMTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo
                dT. Library constructed by Life Technologies.
                Investigator providing samples: Gilbert Smith, NIH"
BASE COUNT   232 a 204 c 266 g 215 t
ORIGIN
Query Match      9.9%; Score 34.6; DB 12; Length 917;
Best Local Similarity 59.8%; Pred. No. 19;
Matches 58; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
QY  152  ACCCCAGCATCACTGTGTATACGCCCAAGAACAAATACCGTGTGATTGATTCAGCTGGAGA 211
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  188  ATCCAAGATTCTCTGGCTTCAGCGGCAAGCACTTCACCAATGAACCTCAGTCATCCGCCA 129
QY  212  AAATTAGAGCCCTTCCAATGTGAGCACAGTCATCCA 248
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  128  TCCTTCAGCCTCTTCCACCAGTGAACCTCAGTCATCCA 92

Search completed: November 6, 2002, 15:26:59
Job time : 2220 secs
```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 04:21:31 ; Search time 801 seconds  
(without alignments)  
155.468 Million cell updates/sec

Title: US09897438BK-2  
Perfect score: 351  
Sequence: 1 GARCARTGYGNACNATNAT.....AYGARGCNTGTGGCNYTN 351

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 320260 seqs, 177392727 residues

Word size : 0

Total number of hits satisfying chosen parameters: 640520

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

- Database : Published\_Applications\_NA:\*
- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
  - 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
  - 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
  - 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
  - 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
  - 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
  - 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
  - 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
  - 9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
  - 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
  - 11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
  - 12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
  - 13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
  - 14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	16	4.6	2379	10	US-09-781-080B-12
2	15	4.3	696	10	US-09-810-052-3
3	15	4.3	729	10	US-09-810-052-6
4	14	4.0	792	10	US-09-590-017-4
5	14	4.0	1584	10	US-09-732-224-3
6	14	4.0	1629	10	US-09-732-224-6
7	14	4.0	1728	10	US-09-794-589-3
8	14	4.0	2094	10	US-09-899-471-6
9	13	3.7	51	10	US-09-853-253-18
10	13	3.7	693	10	US-09-728-911-3
11	13	3.7	696	10	US-09-825-561A-8
12	13	3.7	882	10	US-09-947-744-3
13	13	3.7	906	10	US-09-948-078-3
14	13	3.7	987	10	US-09-732-227-3
15	13	3.7	1644	10	US-09-819-136-7
16	13	3.7	1965	12	US-10-001-632A-3
17	13	3.7	2082	12	US-09-846-996A-2
18	12	3.4	26	10	US-09-986-676A-4
19	12	3.4	429	10	US-09-934-814-6

20	12	3.4	432	9	US-09-929-230-9	Sequence 9, Appli
21	12	3.4	435	10	US-09-923-995-3	Sequence 3, Appli
22	12	3.4	474	9	US-09-929-230-12	Sequence 12, Appli
23	12	3.4	519	10	US-09-745-003-5	Sequence 5, Appli
24	12	3.4	525	10	US-09-934-814-3	Sequence 3, Appli
25	12	3.4	537	10	US-09-745-003-3	Sequence 3, Appli
26	12	3.4	537	10	US-09-934-814-9	Sequence 9, Appli
27	12	3.4	597	10	US-09-801-231-3	Sequence 3, Appli
28	12	3.4	792	10	US-09-934-814-12	Sequence 12, Appli
29	12	3.4	975	10	US-09-741-711-3	Sequence 3, Appli
30	12	3.4	1038	10	US-09-122-383-13	Sequence 13, Appli
31	12	3.4	1059	9	US-09-911-345-3	Sequence 3, Appli
32	12	3.4	1140	12	US-10-003-356-6	Sequence 6, Appli
33	12	3.4	1200	10	US-09-921-823-3	Sequence 3, Appli
34	12	3.4	1245	12	US-10-005-947-3	Sequence 3, Appli
35	12	3.4	1833	10	US-09-921-823-18	Sequence 18, Appli
36	12	3.4	1833	10	US-09-921-823-21	Sequence 21, Appli
37	12	3.4	2022	10	US-09-899-471-3	Sequence 3, Appli
38	12	3.4	2337	10	US-09-731-179-3	Sequence 9, Appli
39	12	3.4	2781	12	US-10-003-356-9	Sequence 3, Appli
40	12	3.4	4527	10	US-09-901-940-3	Sequence 3, Appli
41	11	3.1	24	10	US-09-865-018-13	Sequence 13, Appli
42	11	3.1	252	10	US-09-733-523-3	Sequence 3, Appli
43	11	3.1	333	10	US-09-750-964-3	Sequence 3, Appli
44	11	3.1	426	10	US-09-781-077-6	Sequence 6, Appli
45	11	3.1	435	10	US-09-740-638-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1  
US-09-781-080B-12  
; Sequence 12, Application US/09781080B  
; Patent No. US20020142439A1  
; GENERAL INFORMATION:  
; APPLICANT: HOLLOWAY, JIM  
; APPLICANT: SHEPPARD, PAUL  
; APPLICANT: YAMAMOTO, GAYLE  
; TITLE OF INVENTION: Anti-Angiogenic Intestinal Peptides,  
; FILE REFERENCE: 99-82  
; CURRENT APPLICATION NUMBER: US/09/781.080B  
; CURRENT FILING DATE: 2002-01-30  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 12  
; LENGTH: 2379  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Degenerate sequence  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(2379)  
; OTHER INFORMATION: n = A,T,C or G

Query Match 4.6%; Score 16; DB 10; Length 2379;  
Best Local Similarity 100.0%; Pred. No. 2.8e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 NWSNGGNWSNTGYMGN 138  
Db 1512 NWSNGGNWSNTGYMGN 1527

RESULT 2  
US-09-810-052-3  
; Sequence 3, Application US/09810052  
; Patent No. US20020009775A1  
; GENERAL INFORMATION:  
; APPLICANT: Conklin, Darrell C.  
; APPLICANT: Presnell, Scott R.

```
; TITLE OF INVENTION: HELICAL PROTEIN ZALPHA51
; FILE REFERENCE: 00-24
; CURRENT APPLICATION NUMBER: US/09/810,052
; CURRENT FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/190,410
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/199,443
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 696
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: degenerate sequence
; NAME/KEY: misc_feature
; LOCATION: (1)...(696)
; OTHER INFORMATION: n = A,T,C or G
US-09-810-052-3

Query Match          4.3%; Score 15; DB 10; Length 696;
Best Local Similarity 100.0%; Pred. No. 0.00027;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 257 AYYTNCNGARGARG 271
      |||||||
Db 449 AYYTNCNGARGARG 463

RESULT 3
US-09-810-052-6
; Sequence 6, Application US/09810052
; Patent No. US20020009775A1
; GENERAL INFORMATION:
; APPLICANT: Conklin, Darrell C.
; APPLICANT: Presnell, Scott R.
; TITLE OF INVENTION: HELICAL PROTEIN ZALPHA51
; FILE REFERENCE: 00-24
; CURRENT APPLICATION NUMBER: US/09/810,052
; CURRENT FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/190,410
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/199,443
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 729
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: degenerate sequence
; NAME/KEY: misc_feature
; LOCATION: (1)...(729)
; OTHER INFORMATION: n = A,T,C or G
US-09-810-052-6

Query Match          4.3%; Score 15; DB 10; Length 729;
Best Local Similarity 100.0%; Pred. No. 0.00027;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 257 AYYTNCNGARGARG 271
      |||||||
Db 482 AYYTNCNGARGARG 496

RESULT 4
US-09-990-017-4
; Sequence 4, Application US/09990017
; Patent No. US20020115168A1
; GENERAL INFORMATION:
; APPLICANT: Conklin, Darrell C.
```

```
; APPLICANT: Gao, Zeren
; TITLE OF INVENTION: NOVEL PROTEIN ZLMDA2
; FILE REFERENCE: 00-67
; CURRENT APPLICATION NUMBER: US/09/990,017
; CURRENT FILING DATE: 2000-11-21
; PRIOR APPLICATION NUMBER: US 60/252,374
; PRIOR FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 792
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: degenerate sequence
; NAME/KEY: misc_feature
; LOCATION: 12, 15, 36, 39, 42, 45, 54, 60, 63, 66, 75, 78, 84, 93, 99,
; LOCATION: 102, 105, 108, 111, 114, 117, 120, 123, 126, 144, 147, 153,
; LOCATION: 162, 168, 174, 180, 183, 186, 189, 198, 204, 210, 213, 219,
; LOCATION: 222, 228, 231, 234, 240, 246, 249, 252, 258, 261, 264
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc_feature
; LOCATION: 273, 279, 282, 294, 300, 303, 306, 309, 321, 327, 345, 372,
; LOCATION: 378, 384, 393, 396, 399, 402, 405, 408, 423, 426, 438, 444,
; LOCATION: 447, 450, 468, 474, 483, 486, 498, 504, 510, 513, 516, 525,
; LOCATION: 528, 531, 534, 537, 543, 549, 552, 564, 585, 591, 594
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc_feature
; LOCATION: 597, 618, 630, 648, 651, 654, 657, 660, 663, 678, 696, 699,
; LOCATION: 702, 705, 708, 711, 714, 717, 720, 723, 729, 738, 750, 756,
; LOCATION: 765, 771, 777, 780, 783, 789, 792
; OTHER INFORMATION: n = A,T,C or G
US-09-990-017-4

Query Match          4.0%; Score 14; DB 10; Length 792;
Best Local Similarity 100.0%; Pred. No. 0.0021;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 215 TMNGGNCNCNWSN 228
      |||||||
Db 647 TMNGGNCNCNWSN 660

RESULT 5
US-09-732-224-3
; Sequence 3, Application US/09732224
; Patent No. US20020095022A1
; GENERAL INFORMATION:
; APPLICANT: Holloway, James L.
; TITLE OF INVENTION: Human Secretin-Like Receptor
; FILE REFERENCE: 99-100
; CURRENT APPLICATION NUMBER: US/09/732,224
; CURRENT FILING DATE: 2000-12-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1584
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: This degenerate nucleotide sequence encodes the
; OTHER INFORMATION: amino acid sequence of SEQ ID NO:2.
; NAME/KEY: variation
; LOCATION: (1)...(1584)
; OTHER INFORMATION: N is any nucleotide.
US-09-732-224-3

Query Match          4.0%; Score 14; DB 10; Length 1584;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 GNCNCNGNGARYTN 69
      |||||||
```

Db 374 GNCNMGNARYTN 387

## RESULT 6

US-09-732-224-6

; Sequence 6, Application US/09732224

; Patent No. US20020095022A1

; GENERAL INFORMATION:

; APPLICANT: Holloway, James L.

; TITLE OF INVENTION: Human Secretin-Like Receptor

; FILE REFERENCE: 99-100

; CURRENT APPLICATION NUMBER: US/09/732,224

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 6

; LENGTH: 1629

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: This degenerate nucleotide sequence encodes the

; NAME/KEY: variation

; LOCATION: (1)...(1629)

; OTHER INFORMATION: N is any nucleotide.

US-09-732-224-6

Query Match

Best Local Similarity 4.0%; Score 14; DB 10; Length 1629;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 56 GNCNMGNARYTN 69

Db 374 GNCNMGNARYTN 387

## RESULT 7

US-09-794-589-3

; Sequence 3, Application US/09794589

; Patent No. US20020004224A1

; GENERAL INFORMATION:

; APPLICANT: Sheppard, Paul O.

; TITLE OF INVENTION: KUNITZ DOMAIN POLYPEPTIDE ZKUN8

; FILE REFERENCE: 00-01

; CURRENT APPLICATION NUMBER: US/09/794,589

; CURRENT FILING DATE: 2001-02-27

; PRIOR APPLICATION NUMBER: US 60/186,069

; PRIOR FILING DATE: 2000-02-29

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 3

; LENGTH: 1728

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: degenerate sequence

; NAME/KEY: misc\_feature

; LOCATION: (1)...(1728)

; OTHER INFORMATION: n = A,T,C or G

US-09-794-589-3

Query Match

Best Local Similarity 4.0%; Score 14; DB 10; Length 1728;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 89 CNACNACGNCWSN 102

Db 584 CNACNACGNCWSN 597

## RESULT 8

US-09-899-471-6

; Sequence 6, Application US/09899471

; Patent No. US20020146763A1

; GENERAL INFORMATION:

; APPLICANT: Gao, Zeren

; TITLE OF INVENTION: Murine Cytokine Receptor

; FILE REFERENCE: 00-46

; CURRENT APPLICATION NUMBER: US/09/899,471

; CURRENT FILING DATE: 2001-07-05

; NUMBER OF SEQ ID NOS: 8

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 6

; LENGTH: 2094

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: This degenerate nucleotide sequence encodes the

; NAME/KEY: misc\_feature

; LOCATION: (1)...(2094)

; OTHER INFORMATION: n = A,T,C or G

US-09-899-471-6

Query Match

Best Local Similarity 4.0%; Score 14; DB 10; Length 2094;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 228 NAAyGTWNSNACNG 241

Db 624 NAAyGTWNSNACNG 637

## RESULT 9

US-09-853-253-18

; Sequence 18, Application US/09853253

; Patent No. US20020055156A1

; GENERAL INFORMATION:

; APPLICANT: JASPEERS, STEPHEN

; APPLICANT: SHEPPARD, PAUL

; APPLICANT: DEISHER, THERESA

; APPLICANT: BISHOP, PAUL

; TITLE OF INVENTION: zsig33-like Peptides

; FILE REFERENCE: 00-30

; CURRENT APPLICATION NUMBER: US/09/853,253

; CURRENT FILING DATE: 2001-05-10

; PRIOR APPLICATION NUMBER: 60/203,300

; PRIOR FILING DATE: 2000-05-11

; NUMBER OF SEQ ID NOS: 28

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 18

; LENGTH: 51

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: degenerate sequence

; NAME/KEY: misc\_feature

; LOCATION: (1)...(51)

; OTHER INFORMATION: n = A,T,C or G

US-09-853-253-18

Query Match

Best Local Similarity 3.7%; Score 13; DB 10; Length 51;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 265 GARGARGCNAARG 277

Db 22 GARGARGCNAARG 34

## RESULT 10

US-09-728-911-3

; Sequence 3, Application US/09728911

; Patent No. US20020012669A1

; GENERAL INFORMATION:

; APPLICANT: Presnell, Scott R.

; APPLICANT: Xu, Wenfeng  
; APPLICANT: Kindsvogel, Wayne  
; APPLICANT: Chen, Zhi  
; TITLE OF INVENTION: Human Cytokine Receptor  
; FILE REFERENCE: 99-93  
; CURRENT APPLICATION NUMBER: US/09/728,911  
; CURRENT FILING DATE: 2000-12-01  
; PRIOR APPLICATION NUMBER: US 60/169,049  
; PRIOR FILING DATE: 1999-12-03  
; PRIOR APPLICATION NUMBER: US 60/232,219  
; PRIOR FILING DATE: 2000-09-13  
; PRIOR APPLICATION NUMBER: US 60/244,610  
; PRIOR FILING DATE: 2000-10-31  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 3  
; LENGTH: 693  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: This degenerate nucleotide sequence encodes the  
; OTHER INFORMATION: amino acid sequence of SEQ ID NO:2.  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(693)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-728-911-3

Query Match 3.7%: Score 13; DB 10; Length 693;  
Best Local Similarity 100.0%; Pred. No. 0.017;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 285 NGTNCARTTCAR 297  
| | | | | | | | | |  
Db 99 NGTNCARTTCAR 111

RESULT 11  
US-09-825-561A-8  
; Sequence 8, Application US/09825561A  
; Patent No. US2002013767A1  
; GENERAL INFORMATION:  
; APPLICANT: Sprecher, Cindy A.  
; APPLICANT: West, James W.  
; APPLICANT: Presnell, Scott R.  
; APPLICANT: Holly, Richard D.  
; APPLICANT: Nelson, Andrew J.  
; TITLE OF INVENTION: SOLUBLE ZALPHAL1 CYTOKINE RECEPTORS  
; FILE REFERENCE: 00-22  
; CURRENT APPLICATION NUMBER: US/09/825,561A  
; CURRENT FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/194,731  
; PRIOR FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/222,121  
; PRIOR FILING DATE: 2000-07-28  
; NUMBER OF SEQ ID NOS: 86  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 8  
; LENGTH: 696  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Degenerate polynucleotide sequence of IL-2Rgamma  
; OTHER INFORMATION: polypeptide as shown in SEQ ID NO:4  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(696)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-825-561A-8

Query Match 3.7%: Score 13; DB 10; Length 696;  
Best Local Similarity 100.0%; Pred. No. 0.017;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 82 YTNAAYACNACNA 94  
| | | | | | | | | |  
Db 1 YTNAAYACNACNA 13

RESULT 12  
US-09-947-744-3  
; Sequence 3, Application US/09947744  
; Patent No. US20020146766A1  
; GENERAL INFORMATION:  
; APPLICANT: Holloway, James L.  
; APPLICANT: Lok, Si  
; TITLE OF INVENTION: Human Vomeronsal Receptor-3  
; FILE REFERENCE: 00-82  
; CURRENT APPLICATION NUMBER: US/09/947,744  
; CURRENT FILING DATE: 2001-09-06  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 882  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: This degenerate nucleotide sequence encodes the  
; OTHER INFORMATION: amino acid sequence of SEQ ID NO:2.  
; NAME/KEY: misc\_feature  
; LOCATION: 21, 24, 27, 36, 42, 48, 51, 57, 63, 66, 72, 78, 81, 87, 93,  
; LOCATION: 96, 99, 111, 114, 120, 123, 132, 135, 141, 144, 150, 156,  
; LOCATION: 165, 168, 171, 186, 189, 192, 195, 198, 210, 213, 225, 228,  
; LOCATION: 237, 261, 264, 273, 279, 282, 288, 291, 294, 297, 303  
; OTHER INFORMATION: n = A,T,C or G  
; NAME/KEY: misc\_feature  
; LOCATION: 306, 309, 315, 318, 321, 327, 330, 333, 339, 345, 348, 351,  
; LOCATION: 354, 357, 363, 366, 369, 378, 384, 396, 399, 408, 414, 420,  
; LOCATION: 432, 438, 444, 450, 453, 456, 459, 462, 465, 477, 480, 483,  
; LOCATION: 489, 492, 495, 501, 504, 510, 516, 519, 522, 528, 531  
; OTHER INFORMATION: n = A,T,C or G  
; NAME/KEY: misc\_feature  
; LOCATION: 540, 549, 558, 561, 567, 579, 582, 585, 588, 591, 594, 597,  
; LOCATION: 600, 603, 606, 609, 612, 621, 627, 630, 633, 636, 639, 645,  
; LOCATION: 651, 666, 669, 675, 678, 681, 687, 690, 693, 705, 708, 714,  
; LOCATION: 720, 723, 726, 729, 732, 741, 747, 753, 756, 765, 768  
; OTHER INFORMATION: n = A,T,C or G  
; NAME/KEY: misc\_feature  
; LOCATION: 771, 774, 780, 783, 786, 789, 798, 804, 807, 810, 816, 819,  
; LOCATION: 828, 831, 840, 846, 849, 855, 858, 861, 864, 867, 870, 873,  
; LOCATION: 876  
; OTHER INFORMATION: n = A,T,C or G  
US-09-947-744-3

Query Match 3.7%: Score 13; DB 10; Length 882;  
Best Local Similarity 100.0%; Pred. No. 0.016;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 72 NACNACNTGYTNN 84  
| | | | | | | | | |  
Db 303 NACNACNTGYTNN 315

RESULT 13  
US-09-948-078-3  
; Sequence 3, Application US/09948078  
; Patent No. US20020147308A1  
; GENERAL INFORMATION:  
; APPLICANT: Holloway, James L.  
; APPLICANT: Lok, Si  
; TITLE OF INVENTION: Human Vomeronsal Receptor  
; FILE REFERENCE: 00-73  
; CURRENT APPLICATION NUMBER: US/09/948,078  
; CURRENT FILING DATE: 2001-09-06  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 3

```

; LENGTH: 906
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: This degenerate nucleotide sequence encodes the
; OTHER INFORMATION: amino acid sequence of SEQ ID NO:2.
; NAME/KEY: misc_feature
; LOCATION: 6, 9, 12, 18, 21, 24, 27, 39, 42, 48, 51, 54, 57, 60, 63,
; LOCATION: 66, 69, 75, 78, 81, 84, 93, 96, 99, 108, 111, 117, 120,
; LOCATION: 123, 126, 129, 135, 141, 150, 156, 159, 168, 171, 174, 177,
; LOCATION: 186, 189, 192, 198, 204, 207, 213, 216, 219, 228, 234, 237
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc_feature
; LOCATION: 240, 249, 252, 261, 267, 270, 273, 276, 279, 282, 285, 291,
; LOCATION: 294, 297, 303, 306, 309, 312, 321, 327, 330, 333, 336, 339,
; LOCATION: 345, 348, 354, 360, 372, 375, 384, 387, 393, 396, 399, 402,
; LOCATION: 414, 423, 426, 441, 450, 453, 456, 471, 480, 483, 495
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc_feature
; LOCATION: 498, 507, 510, 513, 516, 531, 537, 540, 543, 546, 552, 555,
; LOCATION: 558, 564, 570, 573, 579, 582, 585, 591, 597, 600, 603, 606,
; LOCATION: 612, 621, 627, 639, 642, 657, 660, 666, 669, 672, 675, 678,
; LOCATION: 681, 684, 693, 696, 699, 705, 711, 717, 720, 723, 726
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc_feature
; LOCATION: 732, 735, 738, 744, 747, 750, 756, 765, 768, 774, 777, 780,
; LOCATION: 789, 795, 798, 801, 804, 810, 813, 816, 819, 825, 828, 837,
; LOCATION: 840, 843, 846, 849, 855, 858, 864, 873, 876, 879, 885, 888,
; LOCATION: 897, 906
; OTHER INFORMATION: n = A,T,C or G
US-09-948-078-3
Query Match 3.7%; Score 13; DB 10; Length 906;
Best Local Similarity 100.0%; Pred. No. 0.016;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 72 NACNACNTGYTGN 84
Db 291 NACNACNTGYTGN 303
|||||
```

```

RESULT 14
US-09-732-227-3
; Sequence 3, Application US/09732227
; Patent No. US20020004228A1
; GENERAL INFORMATION:
; APPLICANT: Holloway, James L.
; TITLE OF INVENTION: zwf1: A Member of the Von Willebrand
; TITLE OF INVENTION: Factor Type A Domain Superfamily
; FILE REFERENCE: 99-99
; CURRENT APPLICATION NUMBER: US/09/732,227
; CURRENT FILING DATE: 2000-12-07
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 987
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: This degenerate nucleotide sequence encodes the
; OTHER INFORMATION: amino acid sequence of SEQ ID NO:2.
; NAME/KEY: variation
; LOCATION: (1)...(987)
; OTHER INFORMATION: N is any nucleotide.
US-09-732-227-3
Query Match 3.7%; Score 13; DB 10; Length 987;
Best Local Similarity 100.0%; Pred. No. 0.016;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 102 NGTNYTNCARTTY 114
Db 675 NGTNYTNCARTTY 687
|||||
```

```

RESULT 15
US-09-819-136-7
; Sequence 7, Application US/09819136
; Patent No. US20020146789A1
; GENERAL INFORMATION:
; APPLICANT: Conklin, Darrell C.
; APPLICANT: Gao, Zeren
; TITLE OF INVENTION: MULTI-DOMAIN PROTEINASE INHIBITOR
; FILE REFERENCE: 00-25
; CURRENT APPLICATION NUMBER: US/09/819,136
; CURRENT FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: US 60/193,642
; PRIOR FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 1644
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: degenerate sequence
; NAME/KEY: misc_feature
; LOCATION: (1)...(1644)
; OTHER INFORMATION: n = A,T,C or G
US-09-819-136-7
Query Match 3.7%; Score 13; DB 10; Length 1644;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 330 NTAYGARGCNTGY 342
Db 1029 NTAYGARGCNTGY 1041
|||||
```

```

Search completed: November 7, 2002, 06:27:12
Job time : 805 secs
```

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 6, 2002, 23:00:45 ; Search time 89 Seconds  
(without alignments)  
437.968 Million cell updates/sec

Title: US-09-897-438B-2

Perfect score: 117

Sequence: 1 EQCGTTHGNATVCFEPCYGP.....FQWKQSLRGEVVEACWAL 117

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 310279 seqs, 166577418 residues

Word size: 1

Total number of hits satisfying chosen parameters: 619741

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+\_p2n.model -DEV=xl  
-Q=/cgn2\_1/USPTO.spool/US09897438/runat\_06112002\_101958\_11894/app\_query.fasta\_1.263  
-DB=Published Applications\_NA -QFMT=fastap -SUFFIX=oligo.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=quality -THR\_MIN=1  
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09897438 -ECGN\_1\_1\_36\_@runat\_06112002\_101958\_11894  
-NCPU=6 -ICPU=3 -NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FCGAPOP=6  
-FCGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Published Applications\_NA:\*  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
117	100.0	351	10	US-09-897-438B-1	Sequence 1, Appli
27	23.1	11580	10	US-09-880-107-3436	Sequence 3436, Ap
18	15.4	413	10	US-09-864-761-11808	Sequence 11808, A
14	12.0	87	10	US-09-864-761-28379	Sequence 28379, A

c 5	13	11.1	214	10	US-09-864-761-31037	Sequence 31037, A
c 6	13	11.1	499	10	US-09-864-761-14497	Sequence 14497, A
c 7	8	6.8	1132	12	US-10-014-927-20	Sequence 20, Appl
c 8	7	6.0	21	10	US-09-897-438B-10	Sequence 10, Appl
c 9	7	6.0	22	10	US-09-897-438B-6	Sequence 6, Appl
c 10	7	6.0	31	10	US-09-894-698-14	Sequence 14, Appl
c 11	7	6.0	238	10	US-09-783-590-2199	Sequence 2199, Ap
c 12	7	6.0	239	10	US-09-960-352-11910	Sequence 11910, A
c 13	7	6.0	365	10	US-09-864-761-15576	Sequence 15576, A
c 14	7	6.0	404	10	US-09-833-381-647	Sequence 647, App
c 15	7	6.0	637	10	US-09-922-217-945	Sequence 945, App
c 16	7	6.0	637	10	US-09-833-263-945	Sequence 945, App
c 17	7	6.0	669	10	US-09-864-761-18014	Sequence 18014, A
c 18	7	6.0	880	10	US-09-770-445-531	Sequence 531, App
c 19	7	6.0	896	10	US-09-908-805B-28	Sequence 28, Appl
c 20	7	6.0	1008	10	US-09-815-242-8010	Sequence 8010, Ap
c 21	7	6.0	1221	10	US-09-974-300-4246	Sequence 4246, Ap
c 22	7	6.0	2000	10	US-09-887-576-870	Sequence 870, App
c 23	7	6.0	2095	10	US-09-764-877-3948	Sequence 3948, Ap
c 24	7	6.0	2343	10	US-09-810-264-23	Sequence 23, Appl
c 25	7	6.0	2346	10	US-09-815-242-7180	Sequence 7180, Ap
c 26	7	6.0	2608	10	US-09-739-254-16	Sequence 16, Appl
c 27	7	6.0	2608	10	US-09-904-615-16	Sequence 16, Appl
c 28	7	6.0	2628	10	US-09-815-242-4377	Sequence 4377, Ap
c 29	7	6.0	2634	10	US-09-815-242-8458	Sequence 8458, Ap
c 30	7	6.0	2711	10	US-09-764-877-3950	Sequence 3950, Ap
c 31	7	6.0	2712	10	US-09-764-877-3949	Sequence 3949, Ap
c 32	7	6.0	3719	10	US-09-917-800A-1705	Sequence 1705, Ap
c 33	7	6.0	13968	10	US-09-764-869-2224	Sequence 2224, Ap
c 34	7	6.0	16854	10	US-09-764-878-217	Sequence 217, App
c 35	7	6.0	19736	12	US-10-014-502-3	Sequence 3, Appl
c 36	7	6.0	98865	10	US-09-770-689A-3	Sequence 3, Appl
c 37	6	5.1	18	10	US-09-854-799-39	Sequence 39, Appl
c 38	6	5.1	18	10	US-09-897-438B-11	Sequence 11, Appl
c 39	6	5.1	80	10	US-09-864-761-21132	Sequence 21132, A
c 40	6	5.1	92	10	US-09-793-306-156	Sequence 156, App
c 41	6	5.1	95	10	US-09-864-761-30980	Sequence 30980, A
c 42	6	5.1	95	10	US-09-878-574-13429	Sequence 13429, A
c 43	6	5.1	107	10	US-09-878-574-6384	Sequence 6384, Ap
c 44	6	5.1	112	10	US-09-864-761-20139	Sequence 20139, A
c 45	6	5.1	119	10	US-09-294-093B-2470	Sequence 2470, Ap

#### ALIGNMENTS

RESULT 1  
US-09-897-438B-1  
; Sequence 1, Application US/09897438B  
; Patent No. US20020137095A1  
; GENERAL INFORMATION:  
; APPLICANT: Mikoshiba, Katsuhiko  
; APPLICANT: Tate, Naoko  
; TITLE OF INVENTION: REELIN PROTEIN CR-50 EPTOPE REGION  
; FILE REFERENCE: 04853-0076-00000  
; CURRENT APPLICATION NUMBER: US/09/897,438B  
; PRIOR FILING DATE: 2001-07-03  
; PRIOR APPLICATION NUMBER: JP 2000-202801  
; PRIOR FILING DATE: 2000-07-04  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 351  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-09-897-438B-1

Alignment Scores:  
Pred. No.: 8.82e-121 Length: 351  
Score: 117.00 Matches: 117  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 10 Gaps: 0

US-09-897-438B-2 (1-117) x US-09-897-438B-1 (1-351)

Qy 1 GluGlnCysGlyThrIleMetHisGlyAsnAlaValThrPheCysGluProTyrGlyPro 20  
 Db 1 GAGCAGTGTGGCACCACATCATGCGCAATGCTGTACCTTCTGTGAGCCGTACGCCCT 60  
 Qy 21 ArgGluLeuThrThrThrCysLeuAsnThrThrThrAlaSerValLeuGlnPheSerIle 40  
 Db 61 CGAGAGCTGACCACACATGCTGAACACAAACAGCATCTGCTCCAGTTTCCATT 120  
 Qy 41 GlySerGlySerCysArgPheSerTyrSerAspProSerIleThrValSerTyrAlaLys 60  
 Db 121 GGGTCAGGATCATGTCGATTAGTTACTCTGACCCCAACATCACTGTGTATGATGCGCAAG 180  
 Qy 61 AsnAsnThrAlaAspTrpIleGlnLeuGluLysIleArgAlaProSerAsnValSerThr 80  
 Db 181 AACATACCGCTGATGGATTACCTGAGGAAATAGAGCCCTTCCATGTGAGCACA 240  
 Qy 81 ValIleHisIleLeuTyrLeuProGluGluAlaLysGlyGluSerValGlnPheGlnTrp 100  
 Db 241 GTCATCCACATCTGTACCTCCCGAGGAGCAAGAGGAGAGCGTGCAGTTCCAGTGG 300  
 Qy 101 LysGlnAspSerLeuArgValGlyGluValTyrGluAlaCysTrpAlaLeu 117  
 Db 301 AACAGGACAGCTGCGAGTGGGTGAGGTGTATGAGGCTGCTGGGCCCTG 351

RESULT 2

US-09-880-107-3436  
 ; Sequence 3436, Application US/09880107  
 ; Patent No. US20020142981A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Horne, Darci T.  
 ; APPLICANT: Vockley, Joseph G.  
 ; APPLICANT: Scherf, Uwe  
 ; APPLICANT: Gene Logic, Inc.  
 ; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer  
 ; EMBL REFERENCE: 44921-5028-WO

CURRENT APPLICATION NUMBER: US/09/880,107  
 CURRENT FILING DATE: 2001-06-14  
 PRIOR APPLICATION NUMBER: US 60/211,379  
 PRIOR FILING DATE: 2000-06-14  
 PRIOR APPLICATION NUMBER: US 60/237,054  
 PRIOR FILING DATE: 2000-10-03

NUMBER OF SEQ ID NOS: 3950  
 SOFTWARE: Patent In Ver. 2.1  
 SEQ ID NO 3436

LENGTH: 11580

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: Genbank Accession No. US20020142981A1 U79716  
 US-09-880-107-3436

Alignment Scores:

Pred. No.:	6,48e-20	Length:	11580
Score:	27.00	Matches:	27
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	23.08%	Indels:	0
DB:	10	Gaps:	0

US-09-897-438B-2 (1-117) x US-09-880-107-3436 (1-11580)

Qy 28 LeuAsnThrThrThrAlaSerValLeuGlnPheSerIleGlySerGlySerCysArgPhe 47  
 Db 941 CTTAATACACACACAGCTTCTGCTCCATTTTCCATTGGGTCAGGTCATGTCGCTTT 1000  
 Qy 48 SerTyrSerAspProSerIle 54  
 Db 1001 AGTTATTACAGCCCGCAGCATC 1021

RESULT 3

US-09-864-761-11808/c  
 ; Sequence 11808, Application US/09864761  
 ; Patent No. US20020048763A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Penn, Sharron G.  
 ; APPLICANT: Hanzel, David R.  
 ; APPLICANT: Chen, Wensheng  
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
 ; FILE REFERENCE: Acomica-X-1  
 ; CURRENT APPLICATION NUMBER: US/09/864,761  
 ; CURRENT FILING DATE: 2001-05-23  
 ; PRIOR APPLICATION NUMBER: US 60/180,312  
 ; PRIOR FILING DATE: 2000-02-04  
 ; PRIOR APPLICATION NUMBER: US 60/207,456  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: US 09/632,366  
 ; PRIOR FILING DATE: 2000-08-03  
 ; PRIOR APPLICATION NUMBER: GB 24263.6  
 ; PRIOR FILING DATE: 2000-10-04  
 ; PRIOR APPLICATION NUMBER: US 60/236,359  
 ; PRIOR FILING DATE: 2000-09-27  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00664  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00665  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00670  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: US 60/234,687  
 ; PRIOR FILING DATE: 2000-09-21  
 ; PRIOR APPLICATION NUMBER: US 09/608,408  
 ; PRIOR FILING DATE: 2000-06-30  
 ; PRIOR APPLICATION NUMBER: US 09/774,203  
 ; PRIOR FILING DATE: 2001-01-29  
 ; NUMBER OF SEQ ID NOS: 49117  
 ; SOFTWARE: Ancomax Sequence Listing Engine vers. 1.1  
 ; SEQ ID NO 11808  
 ; LENGTH: 413  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; OTHER INFORMATION: MAP TO AC000121.1  
 ; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.55  
 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.57  
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.51  
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.44

US-09-864-761-11808

Alignment Scores:

Pred. No.:	2,14e-11	Length:	413
Score:	18.00	Matches:	18
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	15.38%	Indels:	0
DB:	10	Gaps:	0

US-09-897-438B-2 (1-117) x US-09-864-761-11808 (1-413)

QY 6 IleMethHisGlyAsnAlaValThrPheCysGluProTyrglyProArgGluLeu 23  
 |||  
 Db 308 ATTATGCGATGCGCAATGCCGTACCTCTGTGAACCATATGCCACGAGAACTG 255

RESULT 4

```

US-09-864-761-28379/c
; Sequence 28379, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Hanzel, David R.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US 09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 28379
; LENGTH: 87
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC000121.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.55
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.57
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.51
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.44
; OTHER INFORMATION: SWISSPROT HIT: P76008, EVALUE 3.20e+00
; OTHER INFORMATION: NT HIT: g14826977, EVALUE 5.00e-42
US-09-864-761-28379
  
```

Alignment Scores:  
 Pred. No.: 1,21e-07 Length: 87  
 Score: 14.00 Matches: 14

```

Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 11.97% Indels: 0
DB: 10 Gaps: 0

US-09-897-438b-2 (1-117) x US-09-864-761-28379 (1-87)

QY 6 IleMethHisGlyAsnAlaValThrPheCysGluProTyrgly 19
|||
Db 44 ATTATGCGATGCGCAATGCCGTACCTCTGTGAACCATATGCC 3

RESULT 5
US-09-864-761-31037/c
; Sequence 31037, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US 09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 31037
; LENGTH: 214
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC000121.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.62
; OTHER INFORMATION: EST HUMAN HIT: BE938667.1, EVALUE 1.00e-76
; OTHER INFORMATION: NT HIT: g14826977, EVALUE 1.00e-117
; OTHER INFORMATION: SWISSPROT HIT: O14063, EVALUE 1.20e-00
  
```

US-09-864-761-31037

Alignment Scores:  
Pred. No.: 3,72e-06 Length: 214  
Score: 13.00 Matches: 13  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 11.11% Indels: 0  
DB: 10 Gaps: 0

US-09-897-438B-2 (1-117) x US-09-864-761-31037 (1-214)

QY 105 LeuArgValGlyGluValTyrGluAlaCysTrpAlaLeu 117

Db 137 CTCGCTAGGTGAAGTGTATGAAGCCTGCTGGCCCTTA 99

RESULT 6

US-09-864-761-14497/c

; Sequence 14497, Application US/09864761

; Patent No. US20020048763A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

; FILE REFERENCE: Aecm1ca-X-1

; CURRENT APPLICATION NUMBER: US/09/864,761

; CURRENT FILING DATE: 2001-05-23

; PRIOR APPLICATION NUMBER: US 60/180,312

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: US 60/207,456

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/632,366

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: GB 24263,6

; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: PCT/US01/00666

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00667

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00664

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00669

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00665

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00668

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00663

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00662

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00661

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00670

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: US 60/234,687

; PRIOR FILING DATE: 2000-09-21

; PRIOR APPLICATION NUMBER: US 09/608,408

; PRIOR FILING DATE: 2000-06-30

; PRIOR APPLICATION NUMBER: US 09/774,203

; PRIOR FILING DATE: 2001-01-29

; NUMBER OF SEQ ID NOS: 49117

; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1

; SEQ ID NO 14497

; LENGTH: 499

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: MAP TO AC000121.1  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.62  
US-09-864-761-14497

Alignment Scores:  
Pred. No.: 8,53e-06 Length: 499  
Score: 13.00 Matches: 13  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 11.11% Indels: 0  
DB: 10 Gaps: 0

US-09-897-438B-2 (1-117) x US-09-864-761-14497 (1-499)

QY 105 LeuArgValGlyGluValTyrGluAlaCysTrpAlaLeu 117

Db 391 CTCGCTAGGTGAAGTGTATGAAGCCTGCTGGCCCTTA 353

RESULT 7

US-10-014-927-20/c

; Sequence 20, Application US/10014927

; Patent No. US20020115180A1

; GENERAL INFORMATION:

; APPLICANT: Barla, Andrea

; APPLICANT: Lopato, Sergiy

; APPLICANT: Kalyna, Maria

; APPLICANT: Dorner, Silke

; TITLE OF INVENTION: Solice Factor

; FILE REFERENCE: SONN:01305

; CURRENT APPLICATION NUMBER: US/10/014,927

; CURRENT FILING DATE: 2001-10-23

; PRIOR APPLICATION NUMBER: PCT/AT00/00100

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: A 727/99

; PRIOR FILING DATE: 1999-04-23

; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 20

; LENGTH: 1132

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of the artificial sequence: atSRp34/SRI

US-10-014-927-20

Alignment Scores:

Pred. No.: 6.32 Length: 1132  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 6.84% Indels: 0  
DB: 12 Gaps: 0

US-09-897-438B-2 (1-117) x US-10-014-927-20 (1-1132)

QY 76 SerAsnValSerThrValIleHis 83

Db 593 TCCAAATGCTCCACTGTTATCCAT 570

RESULT 8

US-09-897-438B-10

; Sequence 10, Application US/09897438B

; Patent No. US20020137095A1

; GENERAL INFORMATION:

; APPLICANT: Mikoshiba, Katsuhiko

; APPLICANT: Tate, Naoko

; TITLE OF INVENTION: REELIN PROTEIN CR-50 EPTOPE REGION

; FILE REFERENCE: 04853-0076-00000

; CURRENT APPLICATION NUMBER: US/09/897,438B

; CURRENT FILING DATE: 2001-07-03

; PRIOR APPLICATION NUMBER: JP 2000-202801

; PRIOR FILING DATE: 2000-07-04

; NUMBER OF SEQ ID NOS: 11

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 21
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: primer for PCR
US-09-897-438B-10

Alignment Scores:
Pred. No.: 1.61 Length: 21
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.98% Indels: 0
DB: 1.0 Gaps: 0

US-09-897-438B-2 (1-117) x US-09-897-438B-10 (1-21)

Qy 1 GluInCysGlyThrIleMet 7
Db 1 GAGCAGTGTGGCACCATCATG 21

RESULT 9
US-09-897-438B-6/C
; Sequence 6, Application US/09897438B
; Patent No. US20020137095A1
; GENERAL INFORMATION:
; APPLICANT: Mikoshiba, Katsuhiko
; APPLICANT: Tate, Naoko
; TITLE OF INVENTION: REELIN PROTEIN CR-50 EPTOPE REGION
; FILE REFERENCE: 04853-0076-00000
; CURRENT APPLICATION NUMBER: US/09/897.438B
; CURRENT FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: JP 2000-202801
; PRIOR FILING DATE: 2000-07-04
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 22
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: primer for PCR
US-09-897-438B-6

Alignment Scores:
Pred. No.: 1.69 Length: 22
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.98% Indels: 0
DB: 1.0 Gaps: 0

US-09-897-438B-2 (1-117) x US-09-897-438B-6 (1-22)

Qy 111 TyrGluAlaCysTrpAlaLeu 117
Db 21 TATGAGGCGCTGCTGGCCCTG 1

RESULT 10
US-09-894-698-14/C
; Sequence 14, Application US/09894698
; Patent No. US20020026041A1
; GENERAL INFORMATION:
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Wisniewski, Nancy
; TITLE OF INVENTION: FLEA ALLANTOINASE NUCLEIC ACID MOLECULES, PROTEINS AND
; FILE REFERENCE: FC-6-cl-cl
; CURRENT APPLICATION NUMBER: US/09/894.698
```

```
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: 09/543,668
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: 60/128,704
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 31
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Primer
US-09-894-698-14

Alignment Scores:
Pred. No.: 2.36 Length: 31
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.98% Indels: 0
DB: 1.0 Gaps: 0

US-09-897-438B-2 (1-117) x US-09-894-698-14 (1-31)

Qy 36 LeuGlnPheSerIleGlySer 42
Db 23 TTGCAATTGACATAGGATCC 3

RESULT 11
US-09-783-590-2199/C
; Sequence 2199, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783.590
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2199
; LENGTH: 238
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (10)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (22)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (43)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (93)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (117)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (121)
; OTHER INFORMATION: n equals a,t,g, or c
```

NAME/KEY: misc feature  
LOCATION: (183)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (198)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (207)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (210)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (213)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (223)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (224)  
OTHER INFORMATION: n equals a,t,g, or c  
US-09-783-590-2199

Alignment Scores:  
Pred. No.: 17.4 Length: 238  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 5.98% Indels: 0  
DB: 10 Gaps: 0

US-09-897-438B-2 (1-117) x US-09-783-590-2199 (1-238)

QY 17 ProTyrGlyProArgGluLeu 23  
|||||  
DB 154 CCCTATGTCAGAGAGCTT 134

RESULT 12  
US-09-960-352-11910/c  
Sequence 11910, Application US/09960352  
Patent No. US20020137139A1  
GENERAL INFORMATION:  
APPLICANT: Warren, Wesley C.  
APPLICANT: Tao, Nengbing  
APPLICANT: Byatt, John C.  
APPLICANT: Mathalagan, Nagappan  
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
FILE REFERENCE: 16511.006/37-21(10298)C  
CURRENT APPLICATION NUMBER: US/09/960,352  
CURRENT FILING DATE: 2001-09-24  
NUMBER OF SEQ ID NOS: 15112  
SEQ ID NO 11910  
LENGTH: 239  
TYPE: DNA  
ORGANISM: Bos taurus  
OTHER INFORMATION: Clone ID: 51-LIB3057-021-Q1-K1-E4  
US-09-960-352-11910

Alignment Scores:  
Pred. No.: 17.5 Length: 239  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 5.98% Indels: 0  
DB: 10 Gaps: 0

US-09-897-438B-2 (1-117) x US-09-960-352-11910 (1-239)

QY 44 SerCysArgPheSerTyrSer 50  
|||||  
DB 153 TCATGTAGATTACAGTTACAGT 133

RESULT 13  
US-09-864-761-15576  
Sequence 15576, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharron G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aeomica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 15576  
LENGTH: 365  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC009476.2  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.93  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5  
US-09-864-761-15576

Alignment Scores:  
Pred. No.: 26.5 Length: 365  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 5.98% Indels: 0  
DB: 10 Gaps: 0

US-09-897-438B-2 (1-117) x US-09-864-761-15576 (1-365)

QY 67 IleGlnLeuGluLysIleArg 73

```
Db 44 ATTCAGTTAGAAAAATACGT 64
|||||
RESULT 14
US-09-833-381-647/c
; Sequence 647, Application US/09833381
; Patent No. US20020132090A1
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: NO. US20020132090A1el Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5800-119
; CURRENT APPLICATION NUMBER: US/09/833,381
; CURRENT FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 647
; LENGTH: 404
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(404)
; OTHER INFORMATION: n = A,T,C or G
US-09-833-381-647

Alignment Scores:
Pred. No.: 29.3 Length: 404
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.98% Indels: 0
DB: 10 Gaps: 0

US-09-897-438B-2 (1-117) x US-09-833-381-647 (1-404)
Qy 42 SerGlySerCysArgPheSer 48
|||||
Db 202 TCGGGAAGCTGCAGATTTCAGT 182
|||||

RESULT 15
US-09-922-217-945
; Sequence 945, Application US/09922217
; Patent No. US20020076414A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yugu
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.471C13
; CURRENT APPLICATION NUMBER: US/09/922,217
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 945
; LENGTH: 637
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 629
; OTHER INFORMATION: n = A,T,C or G

US-09-922-217-945
Alignment Scores:
Pred. No.: 45.7 Length: 637
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.98% Indels: 0
DB: 10 Gaps: 0

US-09-897-438B-2 (1-117) x US-09-922-217-945 (1-637)
Qy 42 SerGlySerCysArgPheSer 48
|||||
Db 372 TCTGGCAGTTGCCCTTCAGT 392
|||||

Search completed: November 7, 2002, 00:48:12
Job time : 91 secs
```

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 6, 2002, 22:05:15 ; Search time 68 Seconds  
(without alignments)  
527.664 Million cell updates/sec

Title: US-09-897-438B-2

Perfect score: 117

Sequence: 1 EOCGTIMHGNAVTFCEPYGP.....FONKQDSLVRGVEVEACWAL 117

Scoring table: OLIGO

Xgapop 60.0, Xgapext 60.0  
Ygapop 60.0, Ygapext 60.0  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 441362 seqs, 153338381 residues

Word size: 1

Total number of hits satisfying chosen parameters: 878600

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp  
-Q=/cgn2.1/USPTO.spool/US09897438/runat\_06112002\_101956\_11823/app\_query.fasta.1.263  
-DB=Issued\_Patents\_NA -QPMT=fastap -SUFFIX=oligo.rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTPMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09897438 -CGN\_1\_31\_runat\_06112002\_101956\_11823 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAPP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Issued\_Patents\_NA: \*  
1: /cgn2.6/ptodata/1/ina/5A.COMB.seq: \*  
2: /cgn2.6/ptodata/1/ina/5B.COMB.seq: \*  
3: /cgn2.6/ptodata/1/ina/6A.COMB.seq: \*  
4: /cgn2.6/ptodata/1/ina/6B.COMB.seq: \*  
5: /cgn2.6/ptodata/1/ina/PTCUS.COMB.seq: \*  
6: /cgn2.6/ptodata/1/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result	No.	Score	Query Match	Length	DB ID	Description
1	117	100.0	11673	4	US-09-334-220-3	Sequence 3, Appli
2	27	23.1	11580	4	US-09-334-220-4	Sequence 4, Appli
3	7	6.0	31	4	US-09-894-698-14	Sequence 14, Appli
4	7	6.0	896	4	US-09-302-769-28	Sequence 28, Appli
5	7	6.0	1125	4	US-09-134-001C-2718	Sequence 2718, Ap
6	7	6.0	1200	4	US-09-222-938A-6	Sequence 6, Appli
7	7	6.0	1276	4	US-08-858-207A-109	Sequence 109, App
8	7	6.0	1666	4	US-09-221-017B-933	Sequence 933, App
9	7	6.0	1851	4	US-09-608-790-2	Sequence 2, Appli
10	7	6.0	1923	1	US-08-077-939-14	Sequence 14, Appli
11	7	6.0	1923	1	US-08-461-599-14	Sequence 14, Appli
12	7	6.0	1923	1	US-08-461-621-14	Sequence 14, Appli

c 13	7	6.0	1923	1	US-08-465-334-14	Sequence 14, Appli
c 14	7	6.0	1974	1	US-08-077-939-16	Sequence 16, Appli
c 15	7	6.0	1974	1	US-08-461-599-16	Sequence 16, Appli
c 16	7	6.0	1974	1	US-08-461-621-16	Sequence 16, Appli
c 17	7	6.0	1974	1	US-08-465-334-16	Sequence 16, Appli
c 18	7	6.0	2631	1	US-08-785-071A-1	Sequence 1, Appli
c 19	7	6.0	2631	3	US-09-012-872-1	Sequence 3, Appli
c 20	7	6.0	3057	1	US-08-551-459-3	Sequence 14, Appli
c 21	7	6.0	3250	4	US-09-122-1268-14	Sequence 5, Appli
c 22	7	6.0	3466	1	US-08-551-459-5	Sequence 7, Appli
c 23	7	6.0	4526	1	US-09-424-283-7	Sequence 3, Appli
c 24	7	6.0	19736	4	US-09-740-035-3	Sequence 39, Appli
c 25	6	5.1	18	4	US-08-392-459-39	Sequence 39, Appli
c 26	6	5.1	18	5	PCT-US91-08525-39	Sequence 20, Appli
c 27	6	5.1	25	1	US-07-884-811-20	Sequence 20, Appli
c 28	6	5.1	25	1	US-07-885-971-20	Sequence 20, Appli
c 29	6	5.1	25	1	US-08-087-783A-20	Sequence 20, Appli
c 30	6	5.1	25	1	US-08-194-088B-20	Sequence 20, Appli
c 31	6	5.1	25	1	US-08-435-501-19	Sequence 19, Appli
c 32	6	5.1	25	1	US-08-435-764-19	Sequence 19, Appli
c 33	6	5.1	25	1	US-08-792-078-19	Sequence 19, Appli
c 34	6	5.1	25	2	US-08-194-087-20	Sequence 20, Appli
c 35	6	5.1	25	5	PCT-US93-04648-20	Sequence 20, Appli
c 36	6	5.1	25	5	PCT-US93-04717-19	Sequence 19, Appli
c 37	6	5.1	39	4	US-09-262-773-161	Sequence 161, App
c 38	6	5.1	46	4	US-08-961-810-93	Sequence 93, Appli
c 39	6	5.1	46	4	US-08-352-902D-93	Sequence 93, Appli
c 40	6	5.1	50	1	US-08-171-389-375	Sequence 375, App
c 41	6	5.1	50	1	US-08-123-936-375	Sequence 375, App
c 42	6	5.1	50	2	US-08-475-228A-375	Sequence 375, App
c 43	6	5.1	50	3	US-08-482-080A-375	Sequence 375, App
c 44	6	5.1	50	4	US-09-354-947-375	Sequence 375, App
c 45	6	5.1	50	5	PCT-US93-12388-375	Sequence 375, App

#### ALIGNMENTS

RESULT 1

US-09-334-220-3

; Sequence 3, Application US/09334220

; Patent No. 6323177

; GENERAL INFORMATION:

; APPLICANT: St. Jude's Children's Research Hospital

; APPLICANT: Curran, Thomas

; APPLICANT: D'Arcangelo, Gabriella

; TITLE OF INVENTION: INTERACTION OF REELIN WITH VERY LOW

; TITLE OF INVENTION: DENSITY LIPOPROTEIN (VLDL) RECEPTOR FOR SCREENING AND

; TITLE OF INVENTION: THERAPIES

; FILE REFERENCE: 2427/0704

; CURRENT APPLICATION NUMBER: US/09/334, 220

; CURRENT FILING DATE: 1999-06-16

; NUMBER OF SEQ ID NOS: 5

; SOFTWARE: FastSeq\_for-Windows\_Version 3.0

; SEQ ID NO 3

; LENGTH: 11673

; TYPE: DNA

; ORGANISM: Mus musculus

US-09-334-220-3

Alignment Scores:

Pred. No.: 1.53e-116

Score: 117.00

Percent Similarity: 100.00%

Best Local Similarity: 100.00%

Query Match: 100.00%

DB: 4

Length: 11673

Matches: 117

Conservative: 0

Mismatches: 0

Indels: 0

Gaps: 0

US-09-897-438B-2 (1-117) x US-09-334-220-3 (1-11673)

Qy 1 GluGlnCysGlyThrIleMetHisClyAsnAlaValThrPheCysGluProTyrGlyPro 20

Db 970 GAGCAGTGTGGCACCATCATCGATGGCAATGCTGTACCTTCTGTGAGCCGACGGCCCT 1029

```
Qy 21 ArgGluLeuThrThrCysLeuAsnThrThrThrAlaSerValLeuGlnPheSerIle 40
Db 1030 CGAGAGTGCACACACATGCTGAACAACAACAGCATGTCTCCTCCAGTTTCCATT 1089
Qy 41 GlySerGlySerCysArgPheSerTyrSerAspProSerIleThrValSerTyrAlaLys 60
Db 1090 GGGTCAGGATCATGCGATTAGTACTCTACCCACGATCATCTGTGTATAGCCCAAG 1149
Qy 61 AsnAsnThrAlaAspTyrPileGlnLeuGluLysIleArgAlaProSerAsnValSerThr 80
Db 1150 AACATACCGTGATTGATTCAGCTGGAGAAATAGAGCCCTTCCAATGTGAGCACA 1209
Qy 81 ValIleHisIleLeuTyrLeuProGluGluAlaLysGlyGluSerValGlnPheGlnTyr 100
Db 1210 GTCATCCACATCTGTACTCTCCGAGGAGCAAGGAGCGTGCAGTTCAGTGG 1269
Qy 101 LysGlnAspSerLeuArgValGlyGluValTyrGluAlaCysTyrAlaLeu 117
Db 1270 AACAGGACAGCGTGGAGTGGGTGAGGTGTATGAGGCGTCTGGGCCCTG 1320
RESULT 2
US-09-334-220-4
; Sequence 4, Application US/09334220
; Patent No. 6323177
; GENERAL INFORMATION:
; APPLICANT: St. Jude's Children's Research Hospital
; APPLICANT: Curran, Thomas
; APPLICANT: D'Arcangelo, Gabriella
; TITLE OF INVENTION: INTERACTION OF REELIN WITH VERY LOW
; TITLE OF INVENTION: DENSITY LIPOPROTEIN (VLDL) RECEPTOR FOR SCREENING AND
; TITLE OF INVENTION: THERAPIES
; FILE REFERENCE: 2427/06704
; CURRENT APPLICATION NUMBER: US/09/334,220
; CURRENT FILING DATE: 1999-06-16
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 11580
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-334-220-4
Alignment Scores:
Pred. No.: 2,95e-19 Length: 11580
Score: 27.00 Matches: 27
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 23.08% Indels: 0
DB: 4 Gaps: 0
US-09-897-438B-2 (1-117) x US-09-334-220-4 (1-11580)
Qy 28 LeuAsnThrThrThrAlaSerValLeuGlnPheSerIleGlySerGlySerCysArgPhe 47
Db 941 CTTAATACAAACAGCTTCTGTCTCCAAATTTTCCATTTGGTCAGTTCATGTCGCTTT 1000
Qy 48 SerTyrSerAspProSerIle 54
Db 1001 AGTTATTCAGACCCAGCAATC 1021
RESULT 3
US-09-894-698-14/C
; Sequence 14, Application US/09894698
; Patent No. 6469152
; GENERAL INFORMATION:
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Wisniewski, Nancy
; TITLE OF INVENTION: FLEA ALLANTOININASE NUCLEIC ACID MOLECULES, PROTEINS AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: FC-6-C1-C1
; CURRENT APPLICATION NUMBER: US/09/894,698
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: 09/543,668
```

```
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: 60/128,704
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 31
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Primer
US-09-894-698-14
Alignment Scores:
Pred. No.: 3.88 Length: 31
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.98% Indels: 0
DB: 4 Gaps: 0
US-09-897-438B-2 (1-117) x US-09-894-698-14 (1-31)
Qy 36 LeuGlnPheSerIleGlySer 42
Db 23 TTGCATTCAGCATAGGATCC 3
RESULT 4
US-09-302-769-28
; Sequence 28, Application US/09302769
; Patent No. 6323317
; GENERAL INFORMATION:
; APPLICANT: HILTON, Douglas J
; APPLICANT: ALEXANDER, Warren S
; APPLICANT: VINEY, Elizabeth M
; APPLICANT: WILLSON, Tracey A
; APPLICANT: RICHARDSON, Rachael T
; APPLICANT: STARR, Robyn
; APPLICANT: NICHOLSON, Sandra E
; APPLICANT: METCALF, Donald
; APPLICANT: NICOLA, Nicos A
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC AGENTS
; FILE REFERENCE: 10976Z
; CURRENT APPLICATION NUMBER: US/09/302,769
; CURRENT FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 08/962,560
; PRIOR FILING DATE: 1997-10-31
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 28
; LENGTH: 896
; TYPE: DNA
; ORGANISM: Mouse
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (4)...(393)
; NAME/KEY: UNSURE
; LOCATION: (551)
; OTHER INFORMATION: n is unsure
; NAME/KEY: UNSURE
; LOCATION: (651)
; OTHER INFORMATION: n is unsure
US-09-302-769-28
Alignment Scores:
Pred. No.: 102 Length: 896
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.98% Indels: 0
DB: 4 Gaps: 0
```



; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 755 PAGE MILL ROAD  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304-1018  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows  
; SOFTWARE: FastSeq for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/221,017B  
; FILING DATE: 23-DEC-1998  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: Pp1182  
; FILING DATE: 31-DEC-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: Pp1546  
; FILING DATE: 30-JAN-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: Pp2911  
; FILING DATE: 09-APR-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/AU98/01023  
; FILING DATE: 10-DEC-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MONROY, Gladys H.  
; REGISTRATION NUMBER: 32,430  
; REFERENCE/DOCKET NUMBER: 27340-20021.00  
; TELEPHONE: 650-813-5600  
; TELEFAX: 650-494-0792  
; TELEX: 706141  
; INFORMATION FOR SEQ ID NO: 933:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1666 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: circular  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: UNKNOWN  
; ORIGINAL SOURCE:  
; ORGANISM: PORPHYROMONAS GINGIVALIS  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 1...1666  
US-09-221-017B-933

Alignment Scores:  
Pred. No.: 187 Length: 1666  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 5.98% Indels: 0  
DB: 4 Gaps: 0

US-09-897-438B-2 (1-117) x US-09-221-017B-933 (1-1666)

QY 93 GlycUservAlGlnPheGln 99  
|||||  
Db 499 GCGAAAGCGTCCAAATCCAA 519

RESULT 9

US-09-608-790-2/c  
; Sequence 2, Application US/09608790  
; Patent No. 6413737  
; GENERAL INFORMATION:  
; APPLICANT: Cohesion Technologies, Inc.  
; APPLICANT: Olsen, David R.

; APPLICANT: Prior, Jeff  
; APPLICANT: Sehl, Louis C.  
; APPLICANT: Wallace, Donald G.  
; TITLE OF INVENTION: Ecarin Polypeptides, Polynucleotides  
; TITLE OF INVENTION: Encoding Ecarin, and Methods for Use Thereof  
; FILE REFERENCE: C99-002  
; CURRENT APPLICATION NUMBER: US/09/608,790  
; CURRENT FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: 60/143,128  
; PRIOR FILING DATE: 1999-07-09  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 1851  
; TYPE: DNA  
; ORGANISM: E. carinatus leucogaster  
US-09-608-790-2

Alignment Scores:  
Pred. No.: 207 Length: 1851  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 5.98% Indels: 0  
DB: 4 Gaps: 0

US-09-897-438B-2 (1-117) x US-09-608-790-2 (1-1851)

QY 9 GlyAsnAlaValThrPheCys 15

Db 128 GGCATGCGAGTGACTTTTGT 108

RESULT 10

US-08-077-939-14/c  
; Sequence 14, Application US/08077939  
; Patent No. 5521088

; GENERAL INFORMATION:

; APPLICANT: FUJII, Toshio  
; APPLICANT: IWAMATSU, Akihiro  
; APPLICANT: YOSHIMOTO, Hiroyuki  
; APPLICANT: MINETOKI, Toshitaka  
; APPLICANT: BOGAKI, Takayuki  
; APPLICANT: NAGASAWA, Naoshi  
; TITLE OF INVENTION: ALCOHOL ACETYLTRANSFERASE GENES AND USE  
; TITLE OF INVENTION: THEREOF  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/077,939  
; FILING DATE: 18-JUN-1992  
; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 184328/1992  
; FILING DATE: 18-JUN-1992

; APPLICATION DATA:  
; APPLICATION NUMBER: JP 62997/1993  
; FILING DATE: 26-FEB-1993

; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768

; REFERENCE/DOCKET NUMBER: 49441/101 KYPA  
; TELEPHONE: (202)672-5300

```
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1923 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 234..1811
; US-08-077-939-14

Alignment Scores:
Pred. No.: 214 Length: 1923
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.98% Indels: 0
DB: 1 Gaps: 0

US-09-897-438B-2 (1-117) x US-08-077-939-14 (1-1923)

Qy 38 PheSerIleGlySerGlySer 44
|||||
Db 958 TTTTCGATCGGTTCTGGAAGT 938

RESULT 11
US-08-461-599-14/c
; Sequence 14, Application US/08461599
; Patent No. 5658777
; GENERAL INFORMATION:
; APPLICANT: FUJII, Toshio
; APPLICANT: IWAMATSU, Akihiro
; APPLICANT: YOSHIMOTO, Hiroyuki
; APPLICANT: MINETOKI, Toshitaka
; APPLICANT: BOGAKI, Takayuki
; APPLICANT: NAGASAWA, Naoshi
; TITLE OF INVENTION: ALCOHOL ACETYLTRANSFERASE GENES AND USE
; NUMBER OF SEQUENCES: 19
; TITLE OF INVENTION: THEREOF
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,599
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/077,939
; FILING DATE: 18-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 62997/1993
; FILING DATE: 26-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 184328/1992
; FILING DATE: 18-JUN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 49441/105 KYPA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
```

```
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1923 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 234..1811
; US-08-461-599-14

Alignment Scores:
Pred. No.: 214 Length: 1923
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 5.98% Indels: 0
DB: 1 Gaps: 0

US-09-897-438B-2 (1-117) x US-08-461-599-14 (1-1923)

Qy 38 PheSerIleGlySerGlySer 44
|||||
Db 958 TTTTCGATCGGTTCTGGAAGT 938

RESULT 12
US-08-461-621-14/c
; Sequence 14, Application US/08461621
; Patent No. 5686284
; GENERAL INFORMATION:
; APPLICANT: FUJII, Toshio
; APPLICANT: IWAMATSU, Akihiro
; APPLICANT: YOSHIMOTO, Hiroyuki
; APPLICANT: MINETOKI, Toshitaka
; APPLICANT: BOGAKI, Takayuki
; APPLICANT: NAGASAWA, Naoshi
; TITLE OF INVENTION: ALCOHOL ACETYLTRANSFERASE GENES AND USE
; NUMBER OF SEQUENCES: 19
; TITLE OF INVENTION: THEREOF
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,621
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/077,939
; FILING DATE: 18-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 62997/1993
; FILING DATE: 26-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 184328/1992
; FILING DATE: 18-JUN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 49441/103 KYPA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
```



## US-08-077-939-16

Alignment Scores:  
Pred. No.: 220 Length: 1974  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 5.98% Indels: 0  
DB: 1 Gaps: 0

US-09-897-438B-2 (1-117) x US-08-077-939-16 (1-1974)

Qy 38 PheSerIleGlySerGlySer 44  
|||||  
Db 1070 TTTTCGATCGGTTCTGGAAGT 1050

## RESULT 15

US-08-461-599-16/C  
: Sequence 16, Application US/08461599  
: Patent No. 5658777  
: GENERAL INFORMATION:  
: APPLICANT: FUJII, Toshio  
: APPLICANT: IWAMATSU, Akihiro  
: APPLICANT: YOSHIMOTO, Hiroyuki  
: APPLICANT: MINETOKI, Toshitaka  
: APPLICANT: BOGAKI, Takayuki  
: APPLICANT: NAGASAWA, Naoshi  
: TITLE OF INVENTION: ALCOHOL ACETYLTRANSFERASE GENES AND USE  
: NUMBER OF SEQUENCES: 19  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Foley & Lardner  
: STREET: 3000 K Street, N.W., Suite 500  
: CITY: Washington, D.C.  
: COUNTRY: USA  
: ZIP: 20007-5109  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/461,599  
: FILING DATE: 05-JUN-1995  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/077,939  
: FILING DATE: 18-JUN-1993  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: JP 62997/1993  
: FILING DATE: 26-FEB-1993  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: JP 184328/1992  
: FILING DATE: 18-JUN-1992  
: ATTORNEY/AGENT INFORMATION:  
: NAME: BENT, Stephen A.  
: REGISTRATION NUMBER: 29,768  
: REFERENCE/DOCKET NUMBER: 49441/105 KYPA  
: TELEPHONE: (202)672-5300  
: TELEFAX: (202)672-5399  
: TELEX: 904136  
: INFORMATION FOR SEQ ID NO: 16:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 1974 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: double  
: TOPOLOGY: linear  
: FEATURE:  
: NAME/KEY: CDS  
: LOCATION: 346..1923  
US-08-461-599-16

Alignment Scores:  
Pred. No.: 220 Length: 1974  
Score: 7.00 Matches: 7  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 5.98% Indels: 0  
DB: 1 Gaps: 0

US-09-897-438B-2 (1-117) x US-08-461-599-16 (1-1974)

Qy 38 PheSerIleGlySerGlySer 44  
|||||  
Db 1070 TTTTCGATCGGTTCTGGAAGT 1050

Search completed: November 6, 2002, 23:36:50  
Job time : 73 secs





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 6, 2002, 20:53:39 ; Search time 315 Seconds  
(without alignments)  
836.457 Million cell updates/sec

Title: US-09-897-438B-2

Perfect score: 117  
Sequence: 1 EQCGTIMHGNNAVTFCEPYGP.....FQWKQDSLVRGEVVEACWAL 117

Scoring table:  
OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Word size: 1

Total number of hits satisfying chosen parameters: 4368727

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q/cgn2.1/USPTO.spool/US09897438/runat\_06112002\_101955\_11799/app\_query.fasta\_1.263  
-DB=N\_Geneseq\_101002 -OPMT=fastap -SUFFIX=oligo.rng -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCLALIGN=20 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09897438.ecgn\_1.1\_79@runat\_06112002\_101955\_11799 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

N\_Geneseq\_101002.\*  
1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	117	100.0	351	24	ABL40165
2	117	100.0	2745	24	AAD22754
3	117	100.0	11673	24	ABA92603
4	117	100.0	11673	24	AB199284
5	27	23.1	11580	24	ABN96939
6	27	23.1	11580	24	ABA92604
7	27	23.1	11632	23	ABA89484
8	24	20.5	2025	24	AAD22779
9	24	20.5	2274	24	AAD22753
10	18	15.4	413	22	AAK07720
11	18	15.4	413	22	AAK33553
12	18	15.4	413	24	ABS08422
13	14	12.0	87	22	AAK20384
14	14	12.0	87	22	AAK46490
15	14	12.0	87	24	ABS20836
16	8	6.8	1132	21	AAK81900
17	8	6.8	6407	24	ABL32943
18	8	6.8	12117	21	AAK96368
19	7	6.0	21	24	ABL40173
20	7	6.0	22	24	ABL40169
21	7	6.0	31	21	AAC93560
22	7	6.0	145	22	AAI61437
23	7	6.0	220	21	AAC08630
24	7	6.0	279	24	ABN23287
25	7	6.0	327	10	AAK90302
26	7	6.0	327	18	AAI70874
27	7	6.0	327	18	AAI51044
28	7	6.0	327	19	AAV18560
29	7	6.0	327	19	AAV18596
30	7	6.0	327	19	AAV03933
31	7	6.0	327	20	AAH22359
32	7	6.0	327	22	AAH22082
33	7	6.0	354	21	AAK05506
34	7	6.0	365	22	AAK30550
35	7	6.0	365	24	ABS05200
36	7	6.0	371	23	AAI71138
37	7	6.0	387	22	AAI40259
38	7	6.0	409	22	AAI66424
39	7	6.0	411	22	AAI01576
40	7	6.0	411	23	ABL97010
41	7	6.0	415	22	AAI19688
42	7	6.0	425	24	ABN21071
43	7	6.0	446	22	AAK75539
44	7	6.0	451	20	AAK33781
45	7	6.0	453	22	AAK59531

# ALIGNMENTS

RESULT 1  
ABL40165  
ID ABL40165 standard; DNA; 351 BP.  
XX AC ABL40165;  
XX 21-MAY-2002 (first entry)  
XX Mouse reelin protein CR-50 epitope region encoding DNA SEQ ID NO:1.  
DE DE Mouse reelin protein CR-50 epitope region; elucidation; neuron;  
XX KW Mouse; reelin protein CR-50 epitope region; neuroprotective; gene; ds.  
XX KW cerebral disturbance; reelin protein; neuroprotective; gene; ds.  
XX OS Mus musculus.  
XX Key Location/Qualifiers  
XX FT 1..351  
XX FT /\*tag= a

**RESULT 2**

AAD22754  
ID AAD22754 standard; cDNA; 2745 BP.

XX AC  
XX AAD22754;  
XX XX  
XX 26-FEB-2002 (first entry)  
DT DT  
XX XX  
DE Mus musculus truncated reelin cDNA.  
XX XX  
KW Mouse; reelin; F-spondin domain; CR-50 epitope; gene therapy; agyria;  
KW polymicrogyria; ectopic gray matter; ss.  
XX XX  
OS Mus musculus.  
XX XX  
FH Key  
FT CDS  
FT FT Location/Qualifiers  
FT FT 283...2052  
FT FT /\*tag= a  
FT FT /product= "Mouse truncated reelin protein"  
FT FT 283...363  
FT FT /\*tag= b  
FT FT mat\_peptide  
FT FT 364..2049  
FT FT /\*tag= c  
FT FT /product= "Mature truncated reelin protein"  
FT FT misc\_feature  
FT FT 284..849  
FT FT /\*tag= d  
FT FT /note= "Encodes F-spondin domain"  
FT FT misc\_feature  
FT FT 970...1320  
FT FT /\*tag= e  
FT FT /note= "Encodes CR-50 epitope region"

EP1149844-A2.  
PN PN  
PD PD  
XX 31-OCT-2001.  
XX PF 11-APR-2001; 2001EP-030341.  
XX PR 11-APR-2000; 2000JP-0109954.  
XX PA (RIKE ) RIKEN KK.  
PI Mikoshiba K, Tabata H, Nakajima K;  
XX XX  
DR WPI; 2002-019320/03.  
XX P-PSDB; AAEL3606.  
XX PT Novel truncated Reelin protein containing F-spondin domain and CR-50  
PT recognition site of Reelin protein, but not having Reelin repeat site,  
PT useful to treat diseases including agyria due to abnormal neuron  
PT alignment -  
XX PS Claim 10; Page 20-26; 47pp; English.  
XX CC The invention relates to a truncated Reelin protein comprising a  
CC F-spondin domain and a CR-50 recognition site but no reelin protein  
CC repeat site. Reelin is an essential molecule in developing a normal  
CC laminated structure of cerebrum. The truncated reelin protein and its  
CC DNA are useful for treating diseases including agyria, polymicrogyria,  
CC and ectopic gray matter due to abnormal neuronal alignment. Truncated  
CC reelin protein DNA is useful in gene therapy. The present sequence is  
CC a cDNA encoding Mus musculus truncated reelin protein.  
XX SQ Sequence 2745 BP; 661 A; 716 C; 714 G; 654 T; 0 other;

Alignment Scores:  
Pred. No.: Length: 2745  
Score: 117.00 Matches: 117  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 24 Gaps: 0

US-09-897-438B-2 (1-117) x AAD22754 (1-2745)

QY	1	Gluln	CysGlyThrIleMetHisGlyAsnAlaValThrPheCysGluProTyrGlyPro	20			
Db	970	GAGCAGTGTGGCACCACATCATGCGCAATGCTGTCACTTCTGTGAGCGGTACGGCCCT	1029				
QY	21	ArgGluLeuThrThrThrCysLeuAsnThrThrThrAlaSerValLeuGlnPheSerIle	40				
Db	1030	CGAGAGCTGACCACCCACCATGCGCTGAACACACACACAGCATCTGCTCCAGTTTTCATT	1089				
QY	41	GlySerGlySerCysArgPheSerTyrSerAspProSerIleThrValSerTyrAlaLys	60				
Db	1090	GGGTCAGGATCATGTCGATTTAGTTACTGCTGACCCACACATCACTGTGTATAGCCCAAG	1149				
QY	61	AsnAsnThrAlaAspTrpIleGlnLeuGluIysIleArgAlaProSerAsnValSerThr	80				
Db	1150	AACAATACCGCTGATTGGATTTCAGCTGGAGAAAAATTAGAGCCCTTCCAATGTGAGCACA	1209				
QY	81	ValIleHisIleLeuTyrLeuProGluGluAlaLysGlyGluSerValGlnPheGlnTrp	100				
Db	1210	GTATCCACATCTGTACCTCCCGAGAGAGCAAGGGAGAGCGTGCAGTCCAGTGG	1269				
QY	101	LysGlnAspSerLeuArgValGlyGluValTyrGluAlaCysTrpAlaLeu	117				
Db	1270	AAACAGACAGCTGCGAGTGGTGAGGTGTATGAGGCTCTGGGCCCTG	1320				
RESULT 3							
ABA92603							
ID	ABA92603	standard;	cdNA;	11673 BP.			
AC	XX						
CC	ABA92603;						
DT	21-MAR-2002	(first entry)					
XX							
DE	Mouse reelin encoding cdNA SEQ ID NO:3.						
XX							
KW	Mouse; reelin; low density lipoprotein receptor; LDLR; neuroprotective						
KW	extracellular glycoprotein; neotropic; antilipemic; Alzheimer's disease						
KW	neurodegenerative disorder; neuronal regeneration; cognitive function;						
KW	lipid metabolism disease; memory; developmental disorder; gene; ss.						
XX							
OS	Mus musculus.						
XX							
Key	Location/Qualifiers						
FT	283..10668						
FT	/*tag= a						
FT	/product= "mouse reelin"						
PN	US6323177-B1.						
XX							
PD	27-NOV-2001.						
XX							
PF	16-JUN-1999;						
XX							
PR	16-JUN-1999;						
XX							
PA	(SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.						
XX							
PI	Curran T, D'Arcangelo G;						
XX							
DR	WPI; 2002-096596/13.						
DR	P-PSDB; ABH05008.						
XX							
PT	Novel composition useful for screening compounds that modulate Reelin						
PT	binding to low density lipoprotein receptor, comprising an isolated						
PT	reelin polypeptide and low density lipoprotein receptor						
XX							
PS	Example 1; Column 65-74; 45pp; English.						
XX							
CC	The present invention describes a composition (I) comprising an						
CC	isolated reelin protein (II) bound to an isolated low density lipoprotein						
CC	receptor (LDLR) (III). (II) is an extracellular glycoprotein of						
CC	approximately 385 kDa containing a small region of similarity with						
CC	F-spondin at the N terminus, a stretch of positively charged amino						

CC acids at the C terminus, and a series of eight internal repeats of  
CC 330-390 amino acids, each repeat containing two related sub-domains  
CC that flank a pattern of conserved cysteine residues known as an  
CC epidermal growth factor (EGF)-like motif. (I) has neuroprotective,  
CC neurotropic and antilipaeamic activities, and can be used as a modulator  
CC of reelin-LDLR interaction. (I) is useful in screen for compounds that  
CC modulate reelin binding to an LDLR, in an assay system, where the assay  
CC system comprises a microplate array and an automated robotic  
CC microprocessor controlled system for adding and removing reagents to  
CC the microplate array. The compounds identified by the above screening  
CC method are useful as therapeutic agents to provide or alleviate a  
CC diverse spectrum of diseases including neurodegenerative disorders such  
CC as Alzheimer's disease, to facilitate neuronal regeneration after  
CC injury, to prevent or alleviate lipid metabolism diseases, to enhance  
CC cognitive functions and memory or to ameliorate other developmental  
CC disorders. The present sequence encodes mouse (Mus musculus) reelin,  
CC which is used in the exemplification of the present invention.

XX  
SQ Sequence 11673 BP; 2831 A; 2985 C; 2985 G; 2872 T; 0 other;

Alignment Scores:  
Pred. No.: 2,7e-110 Length: 11673  
Score: 117.00 Matches: 117  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 24 Gaps: 0

US-09-897-438B-2 (1-117) x ABA92603 (1-11673)

QY 1 GluGlnCysGlyThrIleMetHisGlyAsnAlaValThrPheCysGluProTyrGlyPro 20  
|||  
DB 970 GAGCAGTGTGGCACCACATCATCGATGGCAATGCTGTACCTTCTGTGAGCCGTACGGCCCT 1029  
|||  
QY 21 ArgGluLeuThrThrCysLeuAsnThrThrThrAlaSerValLeuGlnPheSerIle 40  
|||  
DB 1030 CGAGCTGACCCACACATGCTGTACGCTGTGAGCCGTACGGCCCT 1089  
|||  
QY 41 GlySerGlySerCysArgPheSerTyrSerAspProSerIleThrValSerTyrAlaLys 60  
|||  
DB 1090 GGGTCAGGATCATGTCGATTAGTTACTCTGACCCAGCATCACGTGTGTATACGCCAAG 1149  
|||  
QY 61 AsnAsnThrAlaAspTrpIleGlnLeuGluLysIleArgAlaProSerAsnValSerThr 80  
|||  
DB 1150 AACAAATACCGTGATTGGATTTCAGCTGGAGAAAATAGAGCCCTTCCAATGTGAGCACA 1209  
|||  
QY 81 ValIleHisIleLeuTyrLeuProGluAlaLysGlyGluSerValGlnPheGlnTrp 100  
|||  
DB 1210 GTCATCCACATCTCTGACTCTCCCGAGAGACCAAGGGGAGAGCGTCAGTTCACGTGG 1269  
|||  
QY 101 LysGlnAspSerLeuArgValGlyGluValTyrGluAlaCysTrpAlaLeu 117  
|||  
DB 1270 AAACAGGACACGCTGCGAGTGGGTGAGGTGTATGAGGCCCTGCTGGGCCCTG 1320  
|||

RESULT 4  
ABI99284  
ID ID ABI99284 standard; cDNA: 11673 BP.  
XX  
AC ABI99284;  
XX  
DT 07-MAR-2002 (first entry)  
XX  
DE Mouse ischaemic condition related cDNA sequence SEQ ID NO:128.  
XX  
KW Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;  
KW vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.  
XX  
OS Mus musculus.  
XX  
PN WO200188188-A2.  
XX  
PD 22-NOV-2001.  
XX

```
PF 18-MAY-2001; 2001WO-JP04192.
XX
PR 18-MAY-2000; 2000JP-0145977.
XX
PA (UYNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.
XX
PI Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;
XX WPI; 2002-034733/04.
DR P-PSDB; ABB57065.
XX
XX Examining the ischemic condition (e.g. occlusive ischemia) by measuring
PT expression levels of particular genes defined in the specification or
PT by determining the expression profile of a gene group comprising these
PT genes -
XX
PS Claim 2; Page 362-385; 2690pp; English.
XX
CC The present invention describes a method for examining ischaemic
CC conditions, comprising measuring the expression levels of particular
CC genes (I) in a test sample or determining the expression profile of a
CC gene group in the sample comprising genes selected from (I). The method
CC is useful for examining the ischaemic condition (e.g. compressive
CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring
CC expression levels of particular genes (ABI99202 to ABI99912, encoding
CC the protein sequences in ABB57020 to ABB57374) or by determining the
CC expression profile of a gene group comprising these genes. The
CC expression levels or expression profiles produced by these genes are
CC used as an indicator when screening for ischaemic condition-improving
CC drugs or therapeutics for ischaemic diseases. ABI99913 and ABI99914
CC represent PCR primers for a mouse ischaemic condition related sequence,
CC which are used in the exemplification of the present invention.
XX
SQ Sequence 11673 BP; 2831 A; 2985 C; 2985 G; 2872 T; 0 other;

Alignment Scores:
Pred. No.: 2,7e-110 Length: 11673
Score: 117.00 Matches: 117
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 24 Gaps: 0

US-09-897-438B-2 (1-117) x ABI99284 (1-11673)
QY 1 GluGlnCysGlyThrIleMetHisGlyAsnAlaValThrPheCysGluProTyrGlyPro 20
DB 970 GAGCAGTGTGGCACCACATCATGTCATGGCAATGCTGTCACCTTCTGTGAGCGGTACGGCCT 1029
QY 21 ArgGluLeuThrThrThrCysLeuAsnThrThrThrAlaSerValLeuGlnPheSerIle 40
DB 1030 CGAGAGTGTGACCCACCATGCTGTAACACACACACATGTCCTCCAGTTTTCATT 1089
QY 41 GlySerGlySerCysArgPheSerTyrSerAspProSerIleThrValSerTyrAlaLys 60
DB 1090 GGGTCAGGATCATGTCGATTAGTTACTGTGACCCAGCATCATGTCATAGCCCAAG 1149
QY 61 AsnAsnThrAlaAspTrpIleGlnLeuGluLysIleArgAlaProSerAsnValSerThr 80
DB 1150 AACAAATACCCGTGATTGGATTGAGTTCAGTGGAGAAAATTAGAGCCCTTCCAAATGTGAGCACA 1209
QY 81 ValIleHisIleLeuTyrLeuProGluGluAlaLysGlyGluSerValGlnPheGlnTrp 100
DB 1210 GTCATCCACATCTGTACCTCCCGAGGAGAGCCAAAGGGGAGAGCGTGCGAGTTCCAGTGG 1269
QY 101 LysGlnAspSerLeuArgValGlyGluValTyrGluAlaCysTrpAlaLeu 117
DB 1270 AAACAGCAGACGCTCGCAGTGGGTGAGTGTATGAGGCCCTGCTGGGCCCTG 1320

RESULT 5
ABN96939
ID ABN96939 standard; DNA; 11580 BP.
XX
```

```
AC ABN96939;
XX
XX 13-AUG-2002 (first entry)
XX
XX DE Gene #3437 used to diagnose liver cancer.
XX
XX Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW metastatic liver tumour; cytostatic; expression profile; disease state;
KW disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
XX Homo sapiens.
XX
XX W0200229103-A2.
XX
XX 11-APR-2002.
XX
XX 02-OCT-2001; 2001WO-US30589.
XX
XX 02-OCT-2000; 2000US-237054P.
XX
XX (GENE-) GENE LOGIC INC.
XX
XX Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX WPI; 2002-426119/45.
XX
XX Diagnosing and detecting the progression of liver cancer,
PT hepatocellular carcinoma or metastatic liver tumor in a patient,
PT involves detecting the level of expression of two or more genes, in a
PT liver tissue sample -
XX
XX Claim 1; SEQ ID NO 3437; 298pp; English.
XX
XX The invention relates to a novel method for diagnosing and detecting the
XX progression of liver cancer, hepatocellular carcinoma or metastatic liver
XX tumour in a patient, and differentiating metastatic liver cancer from
XX hepatocellular carcinoma in a patient, involving detecting the level of
XX expression of two or more genes represented in ABN93503-ABN97455 in a
XX tissue sample. The method of the invention has hepatotropic, and
XX cytostatic activity. The method is useful for diagnosing and detecting
XX the progression of liver cancer, hepatocellular carcinoma and metastatic
XX liver carcinoma in a patient. The method is useful for identifying
XX expression profiles which serve as useful diagnostic markers as well as
XX markers that can be used to monitor disease states, disease progression,
XX drug toxicity, drug efficacy and drug metabolism.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 11580 BP; 3014 A; 2696 C; 2753 G; 3116 T; 1 other;

Alignment Scores:
Pred. No.: 2,16e-17 Length: 11580
Score: 27.00 Matches: 27
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 23.08% Indels: 0
DB: 24 Gaps: 0

US-09-897-438B-2 (1-117) x ABN96939 (1-11580)
QY 28 LeuAsnThrThrThrAlaSerValLeuGlnPheSerIleGlySerGlySerCysArgPhe 47
DB 941 CTTAATAACACACACAGCTTCTGTCTCCCAATTTTCCATTTGGTGGTTCATGTCGCTTT 1000
QY 48 SerTyrSerAspProSerIle 54
DB 1001 AGTATTTCAGACCCAGCATC 1021

RESULT 6
ABN92604
ID ABN92604 standard; cDNA; 11580 BP.
XX
```

AC ABA92604;

DT 21-MAR-2002 (first entry)

DE Human reelin encoding cDNA SEQ ID NO:4.

XX

KW Human; reelin; low density lipoprotein receptor; LDLR; neuroprotective;

KW extracellular glycoprotein; nontropic; antilipemic; Alzheimer's disease;

KW neurodegenerative disorder; neuronal regeneration; cognitive function;

KW lipid metabolism disease; memory; developmental disorder; gene; ss.

XX

OS Homo sapiens.

XX

PH Key Location/Qualifiers

FT 176..10558

FT /\*tag= a

FT /product= "human reelin"

XX

PN US6323177-B1.

XX

PD 27-NOV-2001.

XX

PF 16-JUN-1999; 99US-0334220.

XX

PR 16-JUN-1999; 99US-0334220.

XX

PA (SUJUD-) ST JUDE CHILDREN'S RES HOSPITAL.

XX

PI Curran T, D'Arcangelo G;

XX

XX WPI: 2002-096596/13.

DR P-PSDB; ABB05007.

DR

XX

PT Novel composition useful for screening compounds that modulate Reelin

PT binding to low density lipoprotein receptor, comprising an isolated

PT Reelin polypeptide and low density lipoprotein receptor

XX

XX Example 1; Column 75-84; 45pp; English.

PS

XX

CC The present invention describes a composition (I) comprising an

CC isolated reelin protein (II) bound to an isolated low density lipoprotein

CC receptor (LDLR) (III). (II) is an extracellular glycoprotein of

CC approximately 385 kDa containing a small region of similarity with

CC F-spondin at the N terminus, a stretch of positively charged amino

CC acids at the C terminus, and a series of eight internal repeats of

CC 350-390 amino acids, each repeat containing two related sub-domains

CC that flank a pattern of conserved cysteine residues known as an

CC epidermal growth factor (EGF)-like motif. (I) has neuroprotective,

CC nontropic and antilipemic activities, and can be used as a modulator

CC of reelin-LDLR interaction. (I) is useful in screen for compounds that

CC modulate reelin binding to an LDLR, in an assay system, where the assay

CC system comprises a microplate array and an automated robotic

CC microprocessor controlled system for adding and removing reagents to

CC the microplate array. The compounds identified by the above screening

CC method are useful as therapeutic agents to provide or alleviate a

CC diverse spectrum of diseases including neurodegenerative disorders such

CC as Alzheimer's disease, to facilitate neuronal regeneration after

CC injury, to prevent or alleviate lipid metabolism diseases, to enhance

CC cognitive functions and memory or to ameliorate other developmental

CC disorders. The present sequence encodes human reelin, which is used in

CC the exemplification of the present invention.

XX

SQ Sequence 11580 BP; 3014 A; 2696 C; 2753 G; 3116 T; 1 other;

Alignment Scores:

Pred. No.:	2,16e-17	Length:	11580
Score:	27.00	Matches:	27
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	23.08%	Indels:	0
DB:	24	Gaps:	0

US-09-897-438B-2 (1-117) x ABA92604 (1-11580)

QY 28 LeuAsnThrThrThrAlaSerValLeuGlnPheSerIleGlySerGlySerCysArgPhe 47

DB 941 CTTAATAACACACAGCTTCTGCTCCTCAATTTTCCATTTGGTTCAGTTCATGTCGCTTT 1000

QY 48 SerTyrSerAspProSerIle 54

DB 1001 AGTTATTTCAGACCCCGCAGCATC 1021

RESULT 7

AAS89484

ID AAS89484 standard; cDNA; 11632 BP.

XX

AC AAS89484;

XX

DT 13-FEB-2002 (first entry)

XX

DE DNA encoding novel human diagnostic protein #25288.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US08631.

XX

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Drmanac RT, Liu C, Tang YT;

XX

XX WPI: 2001-639362/73.

DR P-PSDB; ABG25297.

DR

XX

CC New isolated polynucleotide and encoded polypeptides, useful in

CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits and to assess

CC biodiversity

XX

PS Claim 1; SEQ ID NO 25288; 103pp; English.

XX

CC The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. AAS64197-AAS94564 represent novel human

CC diagnostic coding sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 11632 BP; 3028 A; 2716 C; 2770 G; 3117 T; 1 other;

Alignment Scores:

Pred. No.:	2.17e-17	Length:	11632
------------	----------	---------	-------

```

Score:                27.00      Matches:                27
Percent Similarity:   100.00%    Conservative:       0
Best Local Similarity: 100.00%    Mismatches:        0
Query Match:         23.08%      Indels:            0
DB:                  23          Gaps:                0

US-09-897-438B-2 (1-117) x AAS89484 (1-11632)

Qy 28 LeuAsnThrThrAlaSerValLeuGlnPheSerIleGlySerGlySerCysArgPhe 47
    |||||||
Db 941 CTTAATACAACAACAGCTTCGTCTCCCAATTTCCATTGGTCAGGTCATGTCGCTTT 1000

Qy 48 SerTyrSerAspProSerIle 54
    |||||||
Db 1001 AGTTATTCAGACCCAGCATC 1021

RESULT 8
AAD22779
ID AAD22779 standard; cDNA; 2025 BP.
XX
AC AAD22779;
XX
DT 26-FEB-2002 (first entry)
XX
DE Xenopus laevis intact reelin cDNA.
XX
KW Frog; reelin; F-spondin domain; CR-50 epitope; gene therapy; agyria;
KW polymicrogyria; ectopic gray matter; ss.
XX
OS Xenopus laevis.
XX
FH Key Location/Qualifiers
FT CDS 157..2025
FT FT /*tag= a
FT FT /product= "Xenopus intact reelin protein"
FT FT /trans_except= (pos:1606..1608, aa:Cys)
FT FT /transl_except= (pos:1639..1641, aa:Cys)
FT FT /note= "CDS does not include stop codon"
FT FT /partial
XX
PN EP1149844-A2.
XX
XX 31-OCT-2001.
XX
PF 11-APR-2001; 2001EP-0303411.
XX
PR 11-APR-2000; 2000JP-0109954.
XX
PA (RIKE ) RIKEN KK.
XX
PI Mikoshiba K, Tabata H, Nakajima K;
XX
WPI; 2002-019320/03.
XX
P-PSDB; AAEL3607.
XX
PT Novel truncated Reelin protein containing F-spondin domain and CR-50
PT recognition site of Reelin protein, but not having Reelin repeat site,
PT useful to treat diseases including agyria due to abnormal neuron
PT alignment
XX
PS Example 1; Fig 1A; 47pp; English.
XX
CC The invention relates to a truncated Reelin protein comprising a
CC F-spondin domain and a CR-50 recognition site but no reelin protein
CC repeat site. Reelin is an essential molecule in developing a normal
CC laminated structure of cerebrum. The truncated reelin protein and its
CC DNA are useful for treating diseases including agyria, polymicrogyria,
CC and ectopic gray matter due to abnormal neuronal alignment. Truncated
CC reelin protein DNA is useful in gene therapy. The present sequence is
CC a cDNA encoding Xenopus laevis intact reelin protein.
XX
SQ Sequence 2025 BP; 541 A; 419 C; 460 G; 604 T; 1 other;
XX

```

---

```

Alignment Scores:
Pred. No.: 5.13e-15      Length: 2025
Score: 24.00            Matches: 24
Percent Similarity: 100.00%    Conservative: 0
Best Local Similarity: 100.00%    Mismatches: 0
Query Match: 20.51%      Indels: 0
DB: 24                  Gaps: 0

US-09-897-438B-2 (1-117) x AAD22779 (1-2025)

Qy 29 AsnThrThrThrAlaSerValLeuGlnPheSerIleGlySerGlySerCysArgPheSer 48
    |||||||
Db 931 AACACAACACTAGGCATCTGTTTTCAGTTCATTTGCGTCAGATCGTCAGGTTTCAGC 990

Qy 49 TyrSerAspPro 52
    |||||||
Db 991 TATTCAGACCCCT 1002

RESULT 9
AAD22753
ID AAD22753 standard; cDNA; 2274 BP.
XX
AC AAD22753;
XX
DT 26-FEB-2002 (first entry)
XX
DE Xenopus laevis truncated reelin cDNA.
XX
KW Frog; reelin; F-spondin domain; CR-50 epitope; gene therapy; agyria;
KW polymicrogyria; ectopic gray matter; ss.
XX
OS Xenopus laevis.
XX
FH Key Location/Qualifiers
FT CDS 157..1455
FT FT /*tag= a
FT FT /product= "X. laevis truncated reelin protein"
FT FT sig_peptide 157..234
FT FT /*tag= b
FT FT mat_peptide 235..1452
FT FT /*tag= c
FT FT misc_feature /product= "Mature truncated reelin protein"
FT FT /*tag= d
FT FT misc_feature /note= "Encodes F-spondin domain"
FT FT /*tag= e
FT FT polyA_signal /note= "Encodes CR-50 epitope region"
FT FT 2254..2259
FT FT /*tag= f
XX
PN EP1149844-A2.
XX
XX 31-OCT-2001.
XX
PF 11-APR-2001; 2001EP-0303411.
XX
PR 11-APR-2000; 2000JP-0109954.
XX
PA (RIKE ) RIKEN KK.
XX
PI Mikoshiba K, Tabata H, Nakajima K;
XX
WPI; 2002-019320/03.
XX
P-PSDB; AAEL3605.
XX
PT Novel truncated Reelin protein containing F-spondin domain and CR-50
PT recognition site of Reelin protein, but not having Reelin repeat site,
PT useful to treat diseases including agyria due to abnormal neuron
PT alignment
XX
PS Claim 8; Page 11-16; 47pp; English.
XX

```

CC The invention relates to a truncated reelin protein comprising a  
CC F-spondin domain and a Cr-50 recognition site but no reelin protein  
CC repeat site. Reelin is an essential molecule in developing a normal  
CC laminated structure of cerebrum. The truncated reelin protein and its  
CC DNA are useful for treating diseases including agyria, polymicrogyria,  
CC and ectopic gray matter due to abnormal neuronal alignment. Truncated  
CC reelin protein DNA is useful in gene therapy. The present sequence is  
CC a cDNA encoding *Xenopus laevis* truncated reelin protein.

XX Sequence 2274 BP; 660 A; 450 C; 469 G; 694 T; 1 other;

Alignment Scores:  
Pred. No.: 5.73e-15 Length: 2274  
Score: 24.00 Matches: 24  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 20.51% Indels: 0  
DB: 24 Gaps: 0

US-09-897-438B-2 (1-117) x AAD22753 (1-2274)

QY 29 AsnThrThrAlaSerValLeuGlnPheSerIleGlySerGlySerCysArgPheSer 48

Db 931 AACACAACTACGGCATCTGTTTGCAGTTTCTATTGGTTCAGGATCGTCAGGTTTCAGC 990

QY 49 TyrSerAspPro 52

Db 991 TATTGAGACCCCT 1002

RESULT 10

AAK07720/C

ID AAK07720 standard; DNA; 413 BP.

XX AAK07720;

XX 05-NOV-2001 (first entry)

XX Human brain expressed single exon probe SEQ ID NO: 7711.  
XX Human; brain expressed exon; gene expression analysis; probe;  
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
KW epilepsy; cancer; ss.

XX Homo sapiens.

XX WO200157275-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00667.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human  
XX brains -  
XX Example 4; SEQ ID NO: 7711; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid

XX probes which are derived from genomic sequences expressed in the human

XX brain. They can be used to measure gene expression in brain cell samples,  
XX

CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is one of the probes of the  
CC invention.

XX Sequence 413 BP; 134 A; 86 C; 67 G; 126 T; 0 other;

Alignment Scores:  
Pred. No.: 1.76e-09 Length: 413  
Score: 18.00 Matches: 18  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 15.38% Indels: 0  
DB: 22 Gaps: 0

US-09-897-438B-2 (1-117) x AAK07720 (1-413)

QY 6 IleMethHisGlyAsnAlaValThrPheCysGluProTyrGlyProArgGluLeu 23

Db 308 ATTATGCATGGCAATGCCGTCACCTTCTGTGACCATATGCCACGAGAACTG 255

RESULT 11

AAK33553/C

ID AAK33553 standard; DNA; 413 BP.

XX AAK33553;

XX 06-NOV-2001 (first entry)

XX Human bone marrow expressed single exon probe SEQ ID NO: 8110.  
XX Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.

XX Homo sapiens.

XX WO200157276-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00668.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488900/53.

XX Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human bone marrow -  
XX Example 4; SEQ ID NO: 8110; 658pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid

XX probes which are derived from genomic sequences expressed in the human  
XX bone marrow. They can be used to measure gene expression in bone marrow  
XX samples, which may enable the improved diagnosis and treatment of cancers  
XX such as lymphoma, leukaemia and myeloma. The present sequence is one of  
XX the probes of the invention.

XX Sequence 413 BP; 134 A; 86 C; 67 G; 126 T; 0 other;

Alignment Scores:  
Pred. No.: 1.76e-09 Length: 413  
Score: 18.00 Matches: 18

Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 15.38% Indels: 0  
 DB: 22 Gaps: 0

US-09-897-438B-2 (1-117) x AAK33553 (1-413)

QY 6 IleMethHisGlyAsnAlaValThrPheCysGluProTyrGlyProArgGluLeu 23  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 308 ATTATGATGCAATGCCGTCTCTGTGAACCATATGCGCCACGAGAACTG 255

RESULT 12  
 ABS08422/c  
 ID ABS08422 standard; DNA; 413 BP.  
 XX  
 AC ABS08422;  
 XX  
 DT 19-AUG-2002 (first entry)  
 XX  
 DE Human genome-derived single exon probe from lung SEQ ID No 8413.  
 XX  
 KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;  
 KW chronic obstructive pulmonary disease; interstitial lung disease;  
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemorrhoidosis;  
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
 KW primary ciliary dyskinesia; pulmonary hypertension;  
 KW hyaline membrane disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200186003-A2.  
 XX  
 PD 15-NOV-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00665.  
 XX  
 PR 04-FEB-2000; 2000US-180312P.  
 PR 26-MAY-2000; 2000US-207456P.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-234687P.  
 PR 27-SEP-2000; 2000US-236359P.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2002-114183/15.  
 XX  
 PT Spatially-addressable set of single exon nucleic acid probes, used to  
 PT measure gene expression in human lung samples -  
 XX  
 PS Claim 1; SEQ ID No 8413; 634pp; English.  
 XX  
 CC The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human lung comprising single exon nucleic acid probes having one of  
 CC 12614 nucleic acid sequences mentioned in the specification, or their  
 CC complements or the 12387 open reading frames derived from the 12614  
 CC probes. Also included are a microarray comprising the novel set of  
 CC probes; the novel set of probes which hybridise at high stringency to a  
 CC nucleic acid expressed in the human lung; measuring gene expression in a  
 CC sample derived from human lung, comprising (a) contacting the array with  
 CC a collection of detectably labeled nucleic acids derived from human lung  
 CC mRNA, and (b) measuring the label detectably bound to each probe of  
 CC the array; identifying exons in a eukaryotic genome, comprising  
 CC (a) algorithmically predicting at least one exon from genomic sequences  
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably  
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,

CC having a fragment identical to the predicted exon, the probe is included  
 CC in the above mentioned microarray; assigning exons to a single gene,  
 CC comprising (a) identifying exons from genomic sequence by the method  
 CC above and (b) measuring the expression of each of the exons in several  
 CC tissues and/or cell types using hybridisation to a single exon  
 CC microarrays having a probe with the exon, where a common pattern of  
 CC expression of the exons in the tissues and/or cell types indicates that  
 CC the exons should be assigned to a single gene; a peptide comprising one  
 CC of 12011 sequences, mentioned in the specification, or encoded by the  
 CC probes/open reading frames (ORF). The probes are used for gene  
 CC expression analysis, and for identifying exons in a gene, particularly  
 CC using human lung derived mRNA and for the study of lung diseases  
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease  
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,  
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary  
 CC haemorrhoidosis, pulmonary histiocytosis, lymphangioleiomyomatosis,  
 CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic  
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension  
 CC and hyaline membrane disease. The present sequence is a single exon  
 CC probe of the invention.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 413 BP; 134 A; 86 C; 67 G; 126 T; 0 other;  
 XX

Alignment Scores:  
 Pred. No.: 1.76e-09 Length: 413  
 Score: 18.00 Matches: 18  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 15.38% Indels: 0  
 DB: 24 Gaps: 0

US-09-897-438B-2 (1-117) x ABS08422 (1-413)

QY 6 IleMethHisGlyAsnAlaValThrPheCysGluProTyrGlyProArgGluLeu 23  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 308 ATTATGATGCAATGCCGTCTCTGTGAACCATATGCGCCACGAGAACTG 255

RESULT 13  
 AAK20384/c  
 ID AAK20384 standard; DNA; 87 BP.  
 XX  
 AC AAK20384;  
 XX  
 DT 05-NOV-2001 (first entry)  
 XX  
 DE Human brain expressed single exon probe SEQ ID No: 20375.  
 XX  
 KW Human; brain expressed exon; gene expression analysis; probe;  
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
 KW epilepsy; cancer; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157275-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00667.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.



XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-483446/52.  
XX Single exon nucleic acid probes for analyzing gene expression in human  
PT brains -  
XX  
XX Example 4; SEQ ID NO: 20375; 650pp + Sequence Listing; English.  
XX  
CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is one of the probes of the  
CC invention.  
XX  
SQ Sequence 87 BP; 21 A; 25 C; 19 G; 22 T; 0 other;  
  
Alignment Scores:  
Pred. No.: 5.38e-06 Length: 87  
Score: 14.00 Matches: 14  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 11.97% Indels: 0  
DB: 22 Gaps: 0  
  
US-09-897-438B-2 (1-117) x AAK46490 (1-87)  
  
Qy 6 IleMethHisGlyAsnAlaValThrPheCysGluProTyrGly 19  
Db 44 ATTATGCATGCGCATGCCGTCCCTCTGTGTAACCATATGGC 3  
  
RESULT 14  
AAK46490/C  
ID AAK46490 standard; DNA; 87 BP.  
XX AC AAK46490;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human bone marrow expressed single exon probe SEQ ID NO: 21047.  
XX  
KW Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200157276-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00668.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-488900/53.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human bone marrow -  
XX  
PS Example 4; SEQ ID NO: 21047; 658pp + Sequence Listing; English.

XX  
CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow  
CC samples, which may enable the improved diagnosis and treatment of cancers  
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of  
CC the probes of the invention.  
XX  
SQ Sequence 87 BP; 21 A; 25 C; 19 G; 22 T; 0 other;  
  
Alignment Scores:  
Pred. No.: 5.38e-06 Length: 87  
Score: 14.00 Matches: 14  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 11.97% Indels: 0  
DB: 22 Gaps: 0  
  
US-09-897-438B-2 (1-117) x AAK46490 (1-87)  
  
Qy 6 IleMethHisGlyAsnAlaValThrPheCysGluProTyrGly 19  
Db 44 ATTATGCATGCGCATGCCGTCCCTCTGTGTAACCATATGGC 3  
  
RESULT 15  
ABS20836/G  
ID ABS20836 standard; DNA; 87 BP.  
XX AC ABS20836;  
XX  
DT 19-AUG-2002 (first entry)  
XX  
DE Human genome-derived single exon probe ORF from lung SEQ ID NO 20827.  
XX  
KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;  
KW chronic obstructive pulmonary disease; interstitial lung disease;  
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
KW tuberosus sclerosis; Gaucher's disease; Niemann-Pick disease;  
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
KW primary ciliary dyskinesia; pulmonary hypertension;  
KW hyaline membrane disease; open reading frame; ORF.  
XX  
OS Homo sapiens.  
XX  
PN WO200186003-A2.  
XX  
PD 15-NOV-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00665.  
XX  
PR 04-FEB-2000; 2000US-180312P.  
PR 26-MAY-2000; 2000US-207456P.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-234687P.  
PR 27-SEP-2000; 2000US-236359P.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2002-114183/15.  
XX  
XX Spatially-addressable set of single exon nucleic acid probes, used to  
PT measure gene expression in human lung samples -  
XX  
XX Claim 4; SEQ ID NO 20827; 634pp; English.  
XX  
XX The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived

CC from human lung comprising single exon nucleic acid probes having one of  
CC 12614 nucleic acid sequences mentioned in the specification, or their  
CC complements or the 12387 open reading frames derived from the 12614  
CC probes. Also included are a microarray comprising the novel set of  
CC probes; the novel set of probes which hybridise at high stringency to a  
CC nucleic acid expressed in the human lung; measuring gene expression in a  
CC sample derived from human lung, comprising (a) contacting the array with  
CC a collection of detectably labeled nucleic acids derived from human lung  
CC mRNA, and (b) measuring the label detectably bound to each probe of  
CC the array; identifying exons in a eukaryotic genome, comprising  
CC (a) algorithmically predicting at least one exon from genomic sequences  
CC of labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
CC having a fragment identical to the predicted exon, the probe is included  
CC in the above mentioned microarray; assigning exons to a single gene,  
CC comprising (a) identifying exons from genomic sequence by the method  
CC above and (b) measuring the expression of each of the exons in several  
CC tissues and/or cell types using hybridisation to a single exon  
CC microarrays having a probe with the exon, where a common pattern of  
CC expression of the exons in the tissues and/or cell types indicates that  
CC the exons should be assigned to a single gene; a peptide comprising one  
CC of 12011 sequences, mentioned in the specification, or encoded by the  
CC probes/open reading frames (ORF). The probes are used for gene  
CC expression analysis, and for identifying exons in a gene, particularly  
CC using human lung derived mRNA and for the study of lung diseases  
CC such as asthma, lung cancer, chronic obstructive pulmonary disease  
CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,  
CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary  
CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,  
CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic  
CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension  
CC and hyaline membrane disease. The present sequence is a single exon  
CC probe open reading frame of the invention.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX  
SQ Sequence 87 BP; 21 A; 25 C; 19 G; 22 T; 0 other;

Alignment Scores:			
Pred. No.:	5.38e-06	Length:	87
Score:	14.00	Matches:	14
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	11.97%	Indels:	0
DB:	24	Gaps:	0

US-09-897-438B-2 (1-117) x ABS20836 (1-87)

Qy	6	IleMetHisGlyAsnAlaValThrPheCysGluProTyrGly	19
Db	44	ATTATGATGCGCAATGCGTCACCTTCTGTGACCATATGCC	3

Search completed: November 6, 2002, 22:12:10  
Job time : 324 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 6, 2002, 20:55:05 ; Search time 2763 Seconds  
(without alignments)  
1232.366 Million cell updates/sec

Title: US-09-897-438B-2

Perfect score: 117

Sequence: 1 EQCGTTHGNNAVTFCEPYGP.....FQWKQDSLRLVGEVVEACWAL 117

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Word size: 1

Total number of hits satisfying chosen parameters: 4106490

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp  
-Q=/cgn2.1/USPTO.spool/US09897438/runat\_06112002\_101956\_11805/app\_query.fasta\_1.263  
-DB=GenEmbl -QFMT=fastap -SUFFIX=oligo.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LISP=45  
-LOCALIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptp  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09897438 -CGN\_1\_1\_2659 -runat\_06112002\_101956\_11805 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAPP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_em.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*

29: em\_vi.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htg\_mus.\*  
34: em\_htg\_pln.\*  
35: em\_htg\_rod.\*  
36: em\_htg\_mam.\*  
37: em\_htg\_vrt.\*  
38: em\_sy.\*  
39: em\_htgo\_hum.\*  
40: em\_htgo\_mus.\*  
41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	117	100.0	11673	6	AX305377 Sequence
2	117	100.0	11673	10	U24703 Mus musculus
c 3	45	38.5	163985	10	AC121878 Mus muscu
5	45	38.5	183156	2	AC023062 Mus muscu
6	27	23.1	11187	10	AB049473 Rattus no
7	27	23.1	11580	6	AX410790 Sequence
8	27	23.1	11580	9	U79716 Human reeli
9	27	23.1	11580	11	G30936 swSS2926 Er
c 10	23	19.7	183641	2	G30938 swSS3176 Er
11	22	18.8	126130	2	AC124933 Rattus no
c 12	22	18.8	202764	2	AC095877 Rattus no
13	18	15.4	10634	5	AC128022 Rattus no
c 14	18	15.4	93163	9	AF090441 Gallus ga
c 15	18	15.4	185996	2	AC001121 Human BAC
16	11	9.4	158	4	AC041023 Homo sapi
17	10	8.5	49858	2	AF232904 Bos tauru
c 18	10	8.5	55700	2	AC008252 Drosophil
19	10	8.5	196594	3	AC017867 Drosophil
20	10	8.5	300412	3	AC010564 Drosophil
c 21	9	7.7	150869	2	AF003471 Drosophil
c 22	9	7.7	153805	2	AF005550 Oryza sat
23	9	7.7	176801	2	AP004661 Oryza sat
c 24	9	7.7	177469	2	AP005493 Oryza sat
c 25	9	7.7	216589	2	AC128345 Rattus no
c 26	9	7.7	217488	9	AC091741 Homo sapi
27	8	6.8	438	11	AP003486 Homo sapi
28	8	6.8	567	9	G02057 STScl9B12 L
c 29	8	6.8	1132	6	AB018508 Homo sapi
c 30	8	6.8	1540	8	AX040663 Sequence
c 31	8	6.8	1765	8	AY128338 Arabidops
c 32	8	6.8	1861	8	AY120722 Arabidops
c 33	8	6.8	2378	8	L26970 Schizosacch
c 34	8	6.8	2632	5	AF050673 Cossypium
c 35	8	6.8	2926	8	GI3753 Chicken mRN
c 36	8	6.8	3361	9	AF173640 Arabidops
c 37	8	6.8	4146	8	AF004169 Pan trogl
c 38	8	6.8	4338	1	AF001035 Arabidops
39	8	6.8	4514	10	U60522 Paracoccus
40	8	6.8	4652	2	AB048843 Mus muscu
c 41	8	6.8	6007	8	AC015122 Drosophil
c 42	8	6.8	6069	8	SCYPL274W
c 43	8	6.8	6407	6	SCYPL276W
c 44	8	6.8	10378	8	AX345845 Sequence
c 45	8	6.8	30544	9	AL117389 S.pombe c
					AL139108 Human DNA

# ALIGNMENTS

RESULT 1

```
AX305377      AX305377      11673 bp      DNA      linear      PAT 11-DEC-2001
LOCUS         Sequence 128 from Patent WO0188188.
DEFINITION    AX305377
ACCESSION     AX305377
VERSION       AX305377.1 GI:17644926
KEYWORDS      house mouse.
SOURCE        Mus musculus
ORGANISM      Mus musculus
REFERENCE     1
AUTHORS       Ishikawa,K., Asai,S., Takahashi,Y., Nagata,T. and Ishii,Y.
TITLE         Method for examining ischemic conditions
JOURNAL       Patent: WO 0188188-A 128 22-NOV-2001;
              School Juridical Person Nihon University (JP)
FEATURES     Location/Qualifiers
              1..11673
               /organism="Mus musculus"
               /db_xref="taxon:10090"
BASE COUNT    2831 a 2985 c 2985 g 2872 t
ORIGIN
Alignment Scores:
Pred. No.:      6.56e-116      Length:      11673
Score:          117.00         Matches:    117
Percent Similarity: 100.00%     Conservative: 0
Best Local Similarity: 100.00%  Mismatches:    0
Query Match:     100.00%       Indels:      0
DB:              6            Gaps:      0

US-09-897-438B-2 (1-117) x AX305377 (1-11673)
Qy 1 GluGlnCysGlyThrIleMetHisGlyAsnAlaValThrPheCysGluProTyrGlyPro 20
|||||
Db 970 GAGCAGTGTGCACCATCATGCGCAATGCTGTCACCTTCTGTGAGCGGTACGGCCCT 1029
|||||
Qy 21 ArgGluLeuThrThrThrCysLeuAsnThrThrThrAlaSerValLeuGlnPheSerIle 40
|||||
Db 1030 CGAGAGCTGACACACCATGCTGCTGAAACACACACATGTCCTCCATTTTCCATT 1089
|||||
Qy 41 GlySerGlySerCysArgPheSerTyrSerAspProSerIleThrValSerTyrAlaLys 60
|||||
Db 1090 GGGTCAGGATCATGCGATTTAGTACTCTGACCCAGCATCATCTGTCATACGCCAAG 1149
|||||
Qy 61 AsnAsnThrAlaAspTrpIleGlnLeuGluLysIleArgAlaProSerAsnValSerThr 80
|||||
Db 1150 AACAAATACCGGTGATTGATTGAGTTCAGTGGAGAAATAGAGCCCTTCCAATGTGAGCACA 1209
|||||
Qy 81 ValIleHisIleLeuTyrLeuProGluGluAlaLysGlyGluSerValGlnPheGlnTrp 100
|||||
Db 1210 GTCATCCACATCCGTGTTACCTCCCGAGGAGCCAAAGGAGAGCGGTGCAAGTTCAGTGG 1269
|||||
Qy 101 LysGlnAspSerLeuArgValGlyGluValTyrGluAlaCysTrpAlaLeu 117
|||||
Db 1270 AAACAGACAGCCCTGCCAGTGGGTGAGTGTATGAGGCCCTGCTGGGCCCTTG 1320
|||||

RESULT 2
MMU24703
LOCUS         Mus musculus reelin mRNA, complete cds.
DEFINITION    MMU24703
ACCESSION     U24703
VERSION       U24703.1 GI:2702252
KEYWORDS      mouse.
SOURCE        Mus musculus
ORGANISM      Mus musculus
REFERENCE     1
AUTHORS       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
              1 (bases 1 to 11673)
              D'Arcangelo,G., Miao,G.G., Chen,S.C., Soares,H.D., Morgan,J.I. and
              Curran,T.
              A protein related to extracellular matrix proteins deleted in the
              mouse mutant reeler
              Nature 374 (6524), 719-723 (1995)
              JOURNAL
```

MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REMARK  
COMMENT  
FEATURES  
source

95231649  
7715726  
2 (bases 1 to 11673)  
D'Arcangelo,G.  
Direct Submission  
Submitted (10-APR-1995) Roche Institute of Molecular Biology, 340  
Kingsland St., Nutley, NJ 07110, USA  
3 (bases 1 to 11673)  
D'Arcangelo,G.  
Direct Submission  
Submitted (19-DEC-1997) Dev. Neurobiology, St. Jude Childr. Res.  
Hosp, 332 N. Lauderdale, Memphis, TN 38105, USA  
Sequence update by submitter  
On Dec 19, 1997 this sequence version replaced gi:902486.  
Location/Qualifiers  
1..11673  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="5"  
/map="D5Gmrl"  
/tissue\_type="cerebellum"  
/dev\_stage="adult"  
1..282  
283..10668  
/codon\_start=1  
/product="reelin"  
/protein\_id="AAB91599.1"  
/db\_xref="GI:2702253"  
/translation="MERGCAWAPRALVLAALLLALTLARAAATCYPRFSPFFFLCTHH  
GELEGGEGEVLISHIAGNPITYPGQEIHVITISTFTFDGLLVLTGLTSTSIQSS  
QSTGGSAFEGIMSDHQFCQPMSCVASHVSHLPTTNLSFVWIAPPAQTCGVNFA  
TATHRGQVIFKDALAQOLCQEQATATAYSHLAEIHSDSVILRDQDSYQLELNP  
IWECSNCEMGEOCGT IMHGNVATFCEPRLTNTCLNTTASVGLQESVQFQKQSL  
SYSDPTVSYAKNTADMIQLEK IRAPSNVSTV IHLXLPPEAKESVQFQKQSL  
RVEGVACWALDNILVINSAREVLEONLDPDVTGNLWLPFGCAVVKHSCQSDGNSI  
YFHNGESFNFAITRDVLDSTEDIQESEESEFQTMGDIILGAVVAGDCGVESGL  
SLVFLKDKRGKJCTPYMDTGYGNLRFYFMVGICDPGVSHENDIILYAKIEGKEHI  
ADLTLYSSYKSVLSVINPELOTPATKCLRQKSHOQYNRNVAWDFHPLVLP  
TSMHMIQFSLNLCGTHQPGNSVLEFSTNHRGWSLLHTECLPCTCAGPLPHSTV  
YSSNYGWNRTITPLPNAALTRDIRWRQTGPIILGNMAIDNVIGPSCILKFCGR  
GQCTRHGCKDCPFGSPACEMASQTFPMFISEFSGARLSYHNFYSIRGAESVFCGG  
VLASGALVFNKDRQLITSLFSDSRFLQTLRLGSKSVLSTLRAPDQPGEGVLL  
HYSYDNGITWKLLEHYSYVNYHEPRIISVELPDPAQFQGFQRMWPQYHSSQGDVMA  
IDEIVMTSLFNSISLDTNLVEVTSGLFVGNQVPCYCHDWTLCFTGDKSLASSR  
YVTCQMIQASVMIQFSLVWGGCQRYTPHMDNQVKLEY SANHGLTWHLVQEECLPSM  
PSCQFTSASIVHASFTQWRVTVLPKPTWSGATFRFWSQSYVTAODPWLADNLYI  
GOOCNMCSCGSHGCHGCRDOGYOCTECHPEALPSTIMSDENPSSWESDQVEI  
GBEVYPECGCGVYSSGSLYFSKAGKRLQVLSWDLTSDWDFVQFTIQIGESAACNK  
PDSREGIILQYSNNGGIOWHLLAEMFSDFSKPRFYVLELPAAGTCTPCTFRFRKPV  
FSGEDYDQWAVDDIILSEKQVQIPVNPPTLPQNYEKPAFDYPMQNSVWLMLANE  
GMAKNSDFCATTPSAMVFGKSDGDFEAVTRDLTKPGYVLQKLTNGCTSQFSTAPV  
LLQYSHDAGMFWLLKEGCFPASAAGCEGNSRELSEPTVYVYTGDFEETRTITAPR  
SLASSKTRFRWTOESSOKNVPFGLDGVYISEPCPYSCGHGDCISGVCFCDLGYTA  
AQGTCVSNTPNISEMDFREGKLSPLWYKITGQVGTGCTLNDGRSLYFNGLGKREA  
RTVPLDTRNISLVQFYIQIGSKTSGITYITPRARYEGLVQVYSDNGLNGLLHLLRELD  
MSFLEPQIISIDLPREAKPATAFRMWQHGKSAQWALGDVILGNSSQTFQDOK  
LDGSDIQANWYRIQGGQVDDICLSMDTALIFTENIGNPRYAEFTDFHVSSEFLQWE  
NMNGCSKPSGAHGIOLQSLNNGKDWLTBECVPTTCGYHTVTSSTVSRFQNW  
RVTVYPLATNSPRTFRWIOINTVAGADSWAIDNVILASGCPMWSGRCIGDSRC  
VCDRGFGFCVPVYVPLPSILKDDFNGLHPLDLPVEVYGAERGNLNGETIKSTCLIF  
GEGRLMLISRLDCTNTWVQFSRLFIAGTPERSHISLLOFSVSGSVYTHLMDFFY  
FPQTTSLFINVPLPAGQATNATFRLMQPYNNGKEEIIIDDFIDGNLNNPILL  
LDTDFGPRDENWFFYPGNGIIGLYCPYSKGAPEDSAMVFSVNEVGEHSITTRDLSV  
NENTIIQFEINVCSTDSSADPVRLEFRSDFGATWHLPLCYLHSSILVSSICSTEH  
HPSYVYACTTQGNRREVFHFKLHLGCSVRFRWQGFYPAGSQPYTWAIDNVYIQGP  
CEMCTGHGSCINGTKICDPTGSGTKISTKNPDFLAKDFEGLEQESDFLMSGGK  
PSRKQGLISSNNLFFNEDLRMLVTRDLDLSHARVQFMRGLGCGKQVDPDRSQVPL  
LQYSLNGLSWLLQEFLEFSNNSVGRYIALSEPLKARSGSLRLRWQVSENGHFSYSP  
WYIDQILIGNISGNTLDFDSTLDSRKWLLHGGTKMPVCGSTGDAIVFTFKASTR  
YVYVTDIAVNESDFIQIDFAASCSTVDSCYALEYSVDLGSWHPLRDCLPTNVFEC  
SYHLQRLIYSDTFNKWTRITLPLSPYSTRSQATRFMRWHPAPDFKQOTWADNVYIGD  
GCLDMCSGHGRCVQSGCVCEQWGGIYCDPEPSTSLPTQLKDNFNRAPSNONMLTVSGG  
KLSTVCGAVASGLAHFSGGCSRLVTVLDNLNTNAEFTIQFYMYGCLTTPSNRNOGVL

LEYSVNGGTTNLLMEIFYDQYKPGFVNILLPPDAKEIATFRFWRQPRHGDQDNDW  
 AIDNVLISGADQRTVMDTFSSAPVPQHERSPADAGPVGRIAPFEMFLEDKTQSVNEMW  
 LFHDDCTVERFCSDPGVMLCGSHDREVIYAVTHDLTPTENWIMQFKISYGCKVPEKI  
 AQOIRVFTSTDFGWSYLVPOCLPADPKSCGSQSPVFFTEGWKRITVPLPESL  
 TGNPFRFYOKYSDVQWADNFVLGPGCLDNGGHCDCLEKQICDPGYSGPNCVLT  
 HSLKTLKREFDSEELKPOLWMSLEGSTCEGVLAENTLYFGSTVQRAOTDOLD  
 LRGAFLQYWGRTGSENNWTSCHRPVCRKEGVLLDFSTGGITWTLLHMDPKYISV  
 RHDYLLPEGALNTTKLRWQPFVINSGLVSVGERAQWALDNILIGGAENIPNSQV  
 TDFDDGSHHEENWSPYNAVRTAGFCNPSFHYLPNKKKDKTHNALSSRELIIQPG  
 YMQFKITVGECEATSCGDLHSVMEYTKDARSWSQVLTQCLPSSNSIGCSPPQFH  
 EATYINAVNSSSKWRTIOLPDHVSSTATOFRIQKQETEKOSWADHWYIGEACPK  
 LCSHGVTGTGAVCIDESQGDSCVSFHELPSYIKDNFESARVTEANNETIOGGVI  
 GSCGQLAPYAHGDSLYFNGCQIRQATKPLDLTRASKINFLQIGSPAGTDCNSDL  
 SGRTVDKAVLLQYVNVNGTTHWIAHQPKDFSTQQRVSYNVPLEARMGVLLRWQ  
 PRNGTGHQDQWALDHVEVLLVSTRKQNYMMNFQRHRLFRNRRSLRRYP"

3'UTR  
 polyA\_signal

BASE COUNT 2831 a 2985 c 2985 g 2872 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 6.56e-116 Length: 11673  
 Score: 117.00 Matches: 117  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 10 Gaps: 0

US-09-897-438b-2 (1-117) x MMU24703 (1-11673)

Qy 1 GluGlnCysGlyThrIleMetHisGlyAsnAlaValThrPheCysGluProTyrGlyPro 20  
 |||||  
 Db 970 GAGCAGTGTGGCACCACATCATCGTGGCAATGCTGTACCTTCTGTGAGCCCTACGGCCCT 1029  
 Qy 21 ArgGluLeuThrThrCysLeuAsnThrThrThrAlaSerValLeuGlnPheSerIle 40  
 |||||  
 Db 1030 CGAGAGCTGACCACACATCGCTGAACAGAACACAGCATCTGTCTCCATTTCCATT 1089  
 Qy 41 GlySerGlySerCysArgPheSerTyrSerAspProSerIleThrValSerTyrAlaLys 60  
 |||||  
 Db 1090 GGGTCAGGATCATGTGCAATTTAGTTACTCTGACCCAGCATCTGTGTATACGCCAAG 1149  
 Qy 61 AsnAsnThrAlaAspTrpIleGlnLeuGluLysIleArgAlaProSerAsnValSerThr 80  
 |||||  
 Db 1150 AACAACTACCGCTGATTGAGTTCAGCTGGAGAAATTAGAGCCCTTCCAAATGTGAGCACA 1209  
 Qy 81 ValIleHisIleLeuTyrLeuProGluGluAlaLysGlyGluSerValGlnPheGlnTrp 100  
 |||||  
 Db 1210 GTCATCCACATCTGTACCTCCCGGAGAACCCAAAGGGAGAGCGTGCACTTCCAGTGG 1269  
 Qy 101 LysGlnAspSerLeuArgValGlyGluValTyrGluAlaCysTrpAlaLeu 117  
 |||||  
 Db 1270 AACAGGACAGCCTCGAGTGGGTGAGGTGTATGAGGCTGCTGGCCCTG 1320

RESULT 3  
 AC121878/c  
 LOCUS AC121878 163985 bp DNA linear ROD 21-JUN-2002  
 DEFINITION Mus musculus clone RP24-131C8, complete sequence.  
 ACCESSION AC121878  
 VERSION AC121878.1 GI:21039993  
 KEYWORDS HTG.  
 SOURCE house mouse.  
 ORGANISM Mus musculus

REFERENCE  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 163985)  
 AUTHORS McPherson, J.D. and Waterston, R.H.  
 TITLE The sequence of Mus musculus clone  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 163985)  
 AUTHORS McPherson, J.D. and Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (21-MAY-2002) Genome Sequencing Center, 4444 Forest Park

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

Parkway, St. Louis, MO 63108, USA  
 3 (bases 1 to 163985)  
 McPherson, J.D. and Waterston, R.H.  
 Direct Submission  
 Submitted (21-JUN-2002) Genome Sequencing Center, 4444 Forest Park  
 Parkway, St. Louis, MO 63108, USA

Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu/gsc/index.shtml  
 Contact: submissions@watson.wustl.edu  
 ----- Project Information -----  
 Center project name: M\_BB0131C08  
 -----

FEATURES  
 source  
 1.163985  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="RP24-131C8"

BASE COUNT 50392 a 33960 c 32422 g 47211 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 4.1e-37 Length: 163985  
 Score: 45.00 Matches: 45  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 38.46% Indels: 0  
 DB: 10 Gaps: 0

US-09-897-438b-2 (1-117) x AC121878 (1-163985)

Qy 73 ArgAlaProSerAsnValSerThrValIleHisIleLeuTyrLeuProGluGluAlaLys 92  
 |||||  
 Db 15382 AGAGCCCTTCCAAATGTGAGCAGCATCATCCATCTGTACCTCCCGAGGAGCAAA 15323  
 Qy 93 GlyGluSerValGlnPheGlnTrpIlyGlnAspSerLeuArgValGlyGluValTyrGlu 112  
 |||||  
 Db 15322 GGGGAGAGCGTGCAGTTCACGTGGAAACAGACAGCCCTGCGAGTGGGTGAGGTGTATGAG 15263

Qy 113 AlaCysTrpAlaLeu 117  
 |||||

Db 15262 GCCTGCTGGGCCCTG 15248

RESULT 4

LOCUS

DEFINITION AC023062 183156 bp DNA linear HTG 05-MAY-2000  
 Mus musculus chromosome 5 clone RP23-466D2 strain C57BL6/J, WORKING  
 DRAFT SEQUENCE, 8 unordered pieces.

ACCESSION AC023062

VERSION AC023062.2 GI:7709916

KEYWORDS HTG; HTGS, PHASE1; HTGS\_DRAFT.

SOURCE Mus musculus.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 183156)

REFERENCE

AUTHORS

Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W.,  
 Bouffard, G.G., Dietrich, N.L., Eagle, W.O., Gupta, J., Ho, S.-L.,  
 Huang, M.C., Idol, J., Lee-Lin, S.-Q., Maduro, O.L., Maduro, V.B.,  
 Mastrian, S.D., McCloskey, J.C., Morse, E., Ojodu, M.A., Pearson, R.,  
 Stantripop, S., Summers, T.J., Thomas, J.W., Thomas, P.J.,  
 Tongson, E.E., Touchman, J.W., Tran, J.T., Vogt, J.L., Walker, M.A.,  
 Wetherby, K.D., and Green, E.D.

TITLE NISC Mouse Sequencing Initiative

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 183156)

AUTHORS

Green, E.D.

Direct Submission

JOURNAL

Submitted (08-FEB-2000) NIH Intramural Sequencing Center, 8717  
 Government Circle, Gaithersburg, MD 20877, USA

COMMENT On May 5, 2000 this sequence version replaced gi:6939141.

----- Genome Center  
Center: NIH Intramural Sequencing Center  
Center code: NISC  
Web site: <http://www.nisc.nih.gov>  
Contact: [nisc.mouse@nhgri.nih.gov](mailto:nisc.mouse@nhgri.nih.gov)  
----- Project Information  
Center project name: Y1  
Center clone name: 466D02

----- Summary Statistics  
Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 175371 bases at least Q40  
Consensus quality: 178474 bases at least Q30  
Consensus quality: 180133 bases at least Q20  
Insert size: 179000; agarose-fp  
Insert size: 188000; pulse-field-gel  
Insert size: 183156; sum-of-contigs  
Quality coverage: 6.96x in Q20 bases; agarose-fp  
Quality coverage: 6.63x in Q20 bases; pulse-field-gel  
Quality coverage: 6.80x in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 8 contigs. The true order of the pieces  
\* is not known and their order in this sequence, record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 2374: contig of 2374 bp in length  
gap of unknown length  
2375 4490: contig of 2116 bp in length  
gap of unknown length  
4491 14959: contig of 10469 bp in length  
gap of unknown length  
14960 31072: contig of 16113 bp in length  
gap of unknown length  
31073 50119: contig of 19047 bp in length  
gap of unknown length  
50120 78196: contig of 28077 bp in length  
gap of unknown length  
78197 109874: contig of 31678 bp in length  
gap of unknown length  
109875 183156: contig of 73282 bp in length.

FEATURES  
source  
1..183156  
/organism="Mus musculus"  
/strain="C57BL6/J"  
/db\_xref="taxon:10090"  
/chromosome="5"  
/clone="Rp23-466D2"  
/clone\_lib="Rpci mouse BAC library 23"  
BASE COUNT 53719 a 37080 c 37302 g 55052 t 3 others  
ORIGIN

Alignment Scores:  
Pred. No.: 4,52e-37 Length: 183156  
Score: 45.00 Matches: 45  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 38,46% Indels: 0  
DB: 2 Gaps: 0

US-09-897-438B-2 (1-117) x AC023062 (1-183156)

Qy 73 ArglaProSerAnValSerThrValIleHisIleLeuTyrLeuProGluGluAlaLys 92  
|||||  
Db 173623 AGACCCCTTCCATGTGAGCACAGTATCATCATCTGTACCTCCCGAGGAGCCAAA 173682  
|||||  
Qy 93 GlyGluSerValGlnPheGlnTrpLysGlnAspSerLeuArgValGlyGluValTyrGlu 112  
|||||  
Db 173683 GGGGAGAGCGTGCAGTTCACAGTGAAGAAACAGACAGCCTGCGAGTGGGTGAGTGATGAG 173742  
|||||

Qy 113 AlaCystTrpAlaLeu 117  
|||||  
Db 173743 GCCTGCTGGCCCTG 173757  
RESULT 5  
AB049473  
LOCUS  
DEFINITION  
Rattus norvegicus mRNA for reelin, complete cds.  
ACCESSION  
AB049473  
VERSION  
AB049473.1 GI:17221617  
SOURCE  
Rattus norvegicus juvenile cerebellum cDNA to mRNA.  
ORGANISM  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE  
1  
Kikkawa,S. and Terashima,T.  
TITLE  
rat reelin (Reln) complete CDS  
JOURNAL  
Published Only in Database (2001)  
REFERENCE  
2 (bases 1 to 11187)  
Kikkawa,S. and Terashima,T.  
AUTHORS  
Direct Submission  
TITLE  
Submitted (29-SEP-2000) Satoshi Kikkawa, Kobe University School of  
Medicine, Department of Anatomy; Chuo-ku, Kusunoki-cho 7-5-1, Kobe,  
Hyogo 650-0017, Japan (E-mail:skikkawa@med.kobe-u.ac.jp,  
Tel:81-78-382-5325, Fax:81-78-382-5328)  
FEATURES  
Location/Qualifiers  
1..11187  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/tissue\_type="cerebellum"  
/dev\_stage="juvenile"  
1..11187  
/gene="Reln"  
/gene="Reln"  
/codon\_start=1  
/product="reelin"  
/protein\_id="BAB78470.1"  
/db\_xref="GI:17221618"  
/translation="MERGWAPRTLVAVLLLLATLRARAATGYPRFSPFLCETH  
HGELEGDQGEVLISLHAGNTLVYPGOEYHVTISTFFDGLVLTGLYTSSTQS  
SOSIGGSAFGFGIMSDHOFQCMCSVVAHSVHSLPTNLSPVWAPAGTSCVNF  
ATATRGQVLEKDALAQCEOGATEATAYSHLHSDSVLRRDDFQSHOLELN  
NIWAECSNDTGEQCGTIIHGNVATECEPGPRELITTYLNTTASVLOFSIGSGCR  
FSYSPSIIVSAKNTADMIQLEIRAPSNVSTIIHLYLPEDAKGNVQFQWQDS  
LHVGEYEAACWALDNLVINSARQVILEDSLDVDTGNMFLFPFATVKHSCQDGA  
IYFPHNEGSQLFATTRDVLSTEDIQEQWSEEFESQPTGWDITLGA VSGECGTIEG  
LSLVFLKDGKRLCTPYMDTGYGNLFYFAMGTCDPGDGHENDVLYAKIEGKEH  
IALDTLSYSYKVPVLVSVVNPLOTPATKCLCRKQNHQGNHNVAVDFRHLPLT  
PTMSHWIQFSLNLCGTHQPCNSVLESTNHGRSWSLLHTECLPEICAGHPHST  
IYSSENISGNRVITFIPNALTRDIRMRQTGPILGNMWAIDNVIIGPSCLKFCG  
RGQTRHKGCKDPFGSPACEMASQTFPMFISFSGSRSLSSYHNFYSTRGAEVSGC  
GVLASGLKFNKDRRLITSLDSSQSFRLQFTLLRGLSKSVLTCRAPDQEGVL  
LHYSDNGITTKLLEHYSYLNHVEPRIISVELPDDAQFGIOFRWQPHSSQEDVW  
AIDEILMTSVLFNSISLDTNLVENVQSLGFLGNIQPCYGHDMTLCFTGDSKLASM  
RYVETCSMOIGASYMIOFSIVMCGCGKYTHMDNVKLEYSTNHGLTWLHWDQCLPS  
MPCQFTSASIHASEFTQWRVRVILFOKTSWGAIRFWSQSYTTAQDENWALDDIY  
IGQCPNMCSDGCHGRCDDQGYGTECHPEAALPTIMSDFNENPNSWSDWDEY  
IGEVVYKPGCGGVSSGSLYFSKAGKRLQVSWDLDTSWDFVQFYIIGGESAACN  
KPDSEEGVLLQYVSNNGIQWHLAEYFSDFKPRFVYLELPAATAKTCTFRFRWQP  
VFSGEDYDQWAVDDIIILSEKOKOVIIPVNPVLTPONFYEKPAFDYPIINOMSVMLAN  
EGMAKNDSFCAVTPSAMVFGKSDGDFAVTRDLTLKPGYVLOFKNLGICASQFSSAP  
VLLOYSHDAGMSFWLVKCGCFPASAGKCGEGNSRELSEPTVITITGDFEWRVITAI  
RLASAKTRFRMIOESSQKNVPFGLDGYIIEPCPSYCSGSGDCCIISGFCFDLGYT  
AAQGTCSNIPNHEMFDEGKLSPLYKISGQVGTGCTGLSDGRSLYFNGLGKRE  
ARTVPLDTRNIRLQVYIQIGSKTSGITCTKPRARNEGLVVQYVSNNDGILWHLRLD  
ELSLFDPOIISIDLPREAKTPATFRWQPHQCKHSAOWALDDVLLGVNDSQTFQD  
KFDGSLDLQANRYRIGQGOVDIDLCSMDTALLFTENIGKPRYAEWDFHVSASSLQF  
DKSMGCSKFPSTHVSQLOYSLNNGKDHHPVTEECVPPTIGCVHVTESSTYSERQCN  
WRVTVYLPATNSPTRFRWQVSNNGKDHHPVTEECVPPTIGCVHVTESSTYSERQCN

CVCDFGPGFPCVVPVPLPSILKDDFNGNLHPDLWPEVYGAERGNLNGETIKSCTSLI  
 YFQPTTLLIFNPVLPYVQFSLRFTAKGTPERSHSILLOSSINGGVYTNKLMDEF  
 LLDFTDFGPREDNWFFYPGGNTGLYCPYSKGAPEEDSAMVFSVNEIGHSTTRDLS  
 VNENTIQIEINVCGSTDDSSADVRLEFRDGFATWHLILLPLCYHSSSLVSLCSTE  
 HHPSSSTYAGTTQWGRVVFHCKGLHLCGSVRFRWQGVFPAGSQPWTADINVTGP  
 QCECMCCGHCSCVNGKICIDPCYSGPTCKISTKPNDFLKDQPEGOLESDELLMSGG  
 KPSKCGIILSSGNLFFNEDGLMLVTRDLDSLHARFVQFFMRGLGCKGVDPDRSPQV  
 LLOYSLNGLSLLEQFELFSNSGVRYIALEMLKARSGSTRLRWQFSENGHYS  
 PWYDILGNGISNTGVLEDFSTLDSRKLWLLHPGTRKMPVCGSDALVFTEKAST  
 RYVWTDIAVNESDFQIDFAASCSTVDSCYAELEYSVDLGSWHLVROCLPTNVE  
 CSRYHLQRLIVLSTNKWTRITLPLPAYTRSOATFRWHPQAPEDKQOTWADINVTIG  
 DCCLMCSHGRCIQSCVDCDEWGLYCDPEPSTLPQOLKNFNAPSNQWLTNG  
 GKUSTVCGAVASGLAIFSGGCSRLVTVDLNLTNAEFTQIFMYICLLITPSSNRNGV  
 LLEYSYNGITLLMEIFYQYSKPFVNIILPPDAKEIGTRFRMRPRHGLDQND  
 WADINLVLSGADQRTVMDLTSSAPQPHERSPADAPGVPIAFIAPMEDKTSVNE  
 WPHDCTVFERFDCDPGVMCLGSHDREVIYAVTDLTPTENWIMQFKISVGCXPEK  
 IAGNTHOVSTDFGVSWSVLPQCLPADPKCSGTVSOPSVPFPTKGWRITVPLPES  
 LAGNVPYFRFYOKYSOVADINFLYLGPGCLNCGGHCGLKEQICIDPCYSGPHCYL  
 THLTKFLERFDESEIKFDLMSLEGGSTCTECGILAENTALIFGGSYVQAITQDL  
 DLRGAKFLQYWGRISENNTSCHRPVCRKEGVLLDYSKDGGITWLLHEMDPQKYS  
 VRHDYTLPEGALTNTLRWQPPFVLSNGLVSGVERAQWALDNLILGAEINLPQ  
 VDTFDEGSGHEENWFYNAVRTAGFCGNPSPHLYPNKKDKTINALSSRLIIOF  
 GYMWOFKIVVGCATSCGDLHSVMLEYTKDARSDSQVLQVQCLPSSNSIGCSPOF  
 HEATYINAVNSSWKRITIQLPDHYSSRATQFRWQGBETEKSQWADIRVYIGERAC  
 RUCSGHGYCTTGAVCIDESFQDDCSVFSHELPSYIKONFESARVTEANWETIQGGA  
 IGSGGOLAPYAHGDSLQNGOIRQAATKPLDLTRASKIMFVLQIGSTAQDTSCNSD  
 LSGPHTVADLVLVYVNGITWHVIAHQPKDFTQAQVSYNVPLKARMKGLVLRW  
 QPRHNGTGHQWALDHVEVVLVSTRKQNMNPSRQGLRHFYNNRRRSRLRRYP"

BASE COUNT 2645 a 3029 c 2930 g 2583 t  
 ORIGIN

## Alignment Scores:

Pred. No.: 1.08e-18 Length: 11187  
 Score: 27.00 Matches: 27  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 23.08% Indels: 0  
 DB: 10 Gaps: 0

US-09-897-438b-2 (1-117) x AB049473 (1-11187)

Qy 28 LeuAsnThrThrAlaSerValLeuGlnPheSerIleGlySerGlySerCysArgPhe 47

Db 1112 CTGAACACGACACAGCATCGTCTCCAGTTTCCATGGTCAGGATCTGTCGATT 1171

Qy 48 SerTyrSerAspProSerIle 54

Db 1172 AGTACTCTGACCCACGATC 1192.

RESULT 6  
 AX410790  
 LOCUS AX410790 11580 bp DNA linear PAT 14-JUN-2002  
 DEFINITION Sequence 3437 from Patent WO0229103.  
 ACCESSION AX410790  
 VERSION AX410790.1 GI:21443495  
 KEYWORDS human.  
 SOURCE Homo sapiens

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1

Auivares.C., Horne.D., Peres-da-Silva,S. and Vockley,J.G.

Gene expression profiles in liver cancer

Patent: WO 0229103-A 3437 11-APR-2002;

GENE LOGIC INC (US)

FEATURES Location/Qualifiers

source

1. .11580

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/note="EMBL/GenBank Accession No. U79716"

BASE COUNT 3014 a 2696 c 2753 g 3116 t 1 others

ORIGIN

## Alignment Scores:

Pred. No.: 1.12e-18 Length: 11580  
 Score: 27.00 Matches: 27  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 23.08% Indels: 0  
 DB: 6 Gaps: 0

US-09-897-438b-2 (1-117) x AX410790 (1-11580)

Qy 28 LeuAsnThrThrAlaSerValLeuGlnPheSerIleGlySerGlySerCysArgPhe 47

Db 941 CTTAATACACACAGCTTCTGCTCCATTTCCATGGTCAGGTTTCATGTCGCTTT 1000

Qy 48 SerTyrSerAspProSerIle 54

Db 1001 AGTATTACACCCACGATC 1021

## RESULT 7

HSU79716

LOCUS Human reelin (RELN) mRNA, complete cds.

ACCESSION U79716

VERSION U79716.1 GI:1743884

KEYWORDS

SOURCE Homo sapiens.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 11580)

AUTHORS DeSilva,U., D'Arcangelo,G., Braden,V.V., Chen,J., Miao,G.G.,

Curran,T. and Green,E.D.

TITLE The human reelin gene: isolation, sequencing, and mapping on

chromosome 7

JOURNAL Genome Res. 7 (2), 157-164 (1997)

MEDLINE 97202106

PUBMED 9049633

REFERENCE 2 (bases 1 to 11580)

AUTHORS DeSilva,U., D'Arcangelo,G., Braden,V.V., Chen,J., Miao,G.G.,

Curran,T. and Green,E.D.

TITLE Direct Submission

JOURNAL Submitted (26-NOV-1996) National Center for Human Genome Research,

National Institutes of Health, 49 Convent Drive, MSC4431, Bethesda,

MD 20892, USA

FEATURES Location/Qualifiers

source

1. .11580

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="7"

gene

1. .11580

/gene="RELN"

/gene="RELN"

/codon\_start=1

/product="reelin"

/protein\_id="AAC51105.1"

/db\_xref="GI:1743885"

/translation="MERSGAROTFLALLLGATLRLARAAAGYVPRFSPFFLCTHHG

ELSGDEGEVLLSHIAGNPTYYVQGVHYVTISTPFDGLLVTLVLTSTSVQASQ

SIGGSAGFGIMSDHQFGNQFMCSSVAVSHVSLPTNLSFIWAIAPAGTCVNFNAT

ATHRGQVIFKDALAQQLCEGAPDVTVHPHLAEIHSDSIILKDDFDSYHQDLQNPNI

WVECNCETEQCGAIHMGNAVTFCEPYGPRELITTLGNTTASVLFQSGSCRSFV

YSDPSITVLYAKNNSADWOLEIRAPSNSVTIHLIYLPEDAKGVQFQWQENLR

VGEVYACWALDNLILNSAHROVLEDSLDVDTGTWMLFFPGCATVHKSCQSDNSIY

FHNGESSENFATRDVLDSTEDIQWSEEFESQPTGMDVLGAVLGTSCGTIESLS

MVFLKDGERRKLCPTSMDTTGYGNLRFYVNGGICDPGNSHENDIILYKIEGKEHIT

LDTLSYSSVPSLVSVINPELQTPATKFLQKQKHQHNRMVAVDFHVLPLVPS

TMSHMIQFSINLCGTHQGNVSLEFSTNHRGSWLLHTLCEIACAGPLPHSTVY

SSENYSQWNRITIPLNAALTRNTRIRWOTGPILGNMWAIDNVIICPSCLKPCSGRG

QCTRHCKCDPGSGPACEMASOTFMFISESGSSRLSSYHNFTYIRCAEVSFGGV

LASGKALVFNKEGRRLITISFLDSSOFLOFTLRGLGSKSVLSTCRAPQPEGVLLH

YSYDNGITWKLLEHYSTLSYHPEIRISLVELPQDAKQFGIQFRWQYHSSQREDVVAI

DEIMTSVFNLSISLDFTNLVEVTQSLGFLVGNVQVPCGHWDTLCTFTGDSKLASSMRY  
VETQSMOIGASVYQISLVNMGCGOKVTPHMDNOVKLEYSNTNHLGTLWHLVOECLPSMP  
SCOEFTSASITVHASERTQWRVVLVLPQKTWSSATRRWSQSYVTAQDEWALDSIYIG  
QCPNMCSCGSDHICRCDQOYQTECHPEAALSTIMDSFENONGWESDQWEVIG  
GEIVPQGGCVIISGSLYFSKAGKRLVSWDLTSDVDFVOFYIIOIGESASCKRP  
DREBGLVLLQYVSNNGIOMHLLAEYFSDSPRFLVLELPAKAACTPCTFRWQCVF  
SGEDYDQAWDDIILISEKQKQIIPVNTLPQNFYKPAFDPNMQMSVFLMANEG  
MVNETFCAPTSAMIFGSDGRFAVTRDLTLKPGYVLOFKLNICANOFSLTAPEVL  
LOYSHDAGSMVFLVKGECYPASAGKCEGNSRELSPTMYHTGDREWRITIVIPRS  
LASKTRFWIOQESSOKNVPFGLQGVYISEPCPSYCSGDCISGECGLDYTA  
QTCVSNVNNHNEFRFEGKLSPLWKITGAQVGGCGTLNDKSLYFNGPKREAR  
TVPLDTRNLRLVYIIOIGESKTSITCIKPRNEGLIVQVSNNDGIMLHLRELDFM  
SFLPEIITDLPQDAPATAPAFWMPQPOKHSQAQWALDDVLIGNMDSQTFQDFK  
DGSIDQANWYRTOGGQVQDIDCLMSDITAFITENIKPRYAEWTFHVSASTFLOEM  
SMGCSKPSNSHVSQVSLNKGMDHLTEECVPTTIGLHVTESSIVTSERFONK  
RTVYLPSTISPTREFRTOQANTVGAOSWADNVVLASGCPMCGRGICDAGCV  
CDRGFGPVCVVPVPLPSILKDFENGLHPLDLPENYVGAERGNLNETIKSGTSLIFK  
GEGRLMISDLQCTNTMVOFSRLFIKASTPERSHILQFISGGITWHLMDPEYF  
PQTNILFNVPLPYTAQTNATFRWQVNNGKKEIWIWDDFIIDGNVNNPVML  
DTDFGPRDNMFYFGNIGLYCPYSSKGAPEDSAMVFSVEGHSITTRDLNVN  
ENTIIQFEINVGSTSSADPVRLFSRDFGATWHLPLCYHSSSHVSLCSTEH  
PSSTYAGTMQGRREVHFKLHLCGSVFRWYQGYFAGSQPVTWADNVYIGQC  
EEMNGGSCINGTKICDPYSGPTCKISTKNPDLKDDFEGQLSDRELLMSGKP  
SRKGLSSGNLFFNEDGLRMLTRDLDLSHAFVOFFEMRLCGCGKVPDPSPQVLL  
QYSLNGLSLWLLQEEFLSNSNVGRYIALEIPLKARSGSTRLMWQPSENGHFYSPW  
VSDTIVAGNEDSLQIDFAASCSTDSVAIELEYSDVLGSLWHLVRLDCLPTNVECS  
RYHLRLVSDTFNKTRITVLPPLVYTRSOATFRWHQAPDFKQOTWALDNYVIGDG  
CIDMCSHGRCIOGNCVDEOMGLXCDDPETSLPQLKDNENRAPSNNWLVNCGCK  
LSVTCVAVSMAHLSGCGSLRLVTLVNLNAEFIQFIYFYGCLITPNNRNQGVLL  
EYVNGIYWNLMELIFDYQSKPGEVNIPLPDPAKEIATFRWQPRHGDLDNDWA  
IDNVLLSGADQVMTLDTFSSAPVQHERSPADAGPVGRIAFDMEDKTSVNEHL  
FHDCTVERFSDPDCVGLCGSHDGEVAVTHDLTPTGECWMOFKISVGCYKSEKIA  
QNOIHVOYSTDFGVSNVYLPOCLPADPKCSGVSPSPVFFPTKWKRTVPLPESLV  
GNVPRFXYQKYSDDMOVADINFEVLPGCLDNCRGHCLREQCICDPYSGPNCYLTH  
TLKTFIKERFDEIEKIDPLWMSLEGGSTCEGILAEIDTALFGGTVRQAVTQDLDL  
RGAKFIVGISENNMTSCHRIKREKGVLLDYSTDGGITWTLHEMDYQKYSVR  
HDYILPEBALNTLRWMPQVVISGVVERAOWALDNILIGGAEINISQLVD  
TFDEGTSHENMSFYPAVRTAGFCGNPSFHLVWPNKKDKTHNALSSRELIIOPGY  
MMQFKIVGCEATSCGDLHSLVMEYTKDARSQWOLVOTQCLPSSNSIGCSPFOFHE  
ATLYNSVSSMKRITVLPDHRVSSATOFRTWIKGEETEKOAWIDHVVIGEACPKL  
CSGHGYCTTGATCIDESEQDDCSVESHDLPSYIKDNFESARTEANNETQGGVIG  
SCGQLAPYAHGDSLFNGCQIQAAATKPLDUTRASKIMFVIOIGMSQTDSCSNDLS  
GPHAVDKVLLQYSVNNGITHVIAHQKDFQAOQVSYNVPLEARMKGLLRWMP  
RHNGTGDQWALDHVLEVLVSTRKQNMNFRSQHGLRHFYNNRRSLRRYP"

BASE COUNT 3014 a 2696 g 3116 t 1 others  
ORIGIN

Alignment Scores:  
Pred. No.: 1.12e-18 Length: 11580  
Score: 27.00 Matches: 27  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 23.08% Indels: 0  
DB: 9 Gaps: 0

US-09-897-438B-2 (1-117) x HSU79716 (1-11580)

Qy 28 LeuAsnThrThrAlaSerValLeuGlnPheSerIleGlySerGlySerCysArgPhe 47  
|||||  
Db 941 CTTAATACACACAGAGCTTCTGCTCCCAATTCCTAGGTCATGTCGCTTT 1000  
|||||  
Qy 48 SerTyrSerAspProSerIle 54  
|||||  
Db 1001 AGTTATTACAGACCCAGCATC 1021  
|||||

RESULT 8  
G30936  
LOCUS  
DEFINITION SWSS2926 Eric D. Green Homo sapiens STS genomic, sequence tagged  
site  
ACCESSION G30936  
VERSION G30936.1 GI:1923209

KEYWORDS STS.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 11580)  
AUTHORS Bouffard,G., Iyer,L.M., Idol,J.R., Braden,V.V., Cunningham,A.F.,  
Weintraub,L.A., Mohr-Tidwell,R.M., Peluso,D.C., Fulton,R.S.,  
Leckie,M.P. and Green,E.D.  
TITLE A collection of 1614 human chromosome 7-specific STSs  
JOURNAL Genome Res. 7 (1), 59-64 (1997)  
MEDLINE 97189344  
PUBMED 9037602  
REFERENCE 2 (bases 1 to 11580)  
AUTHORS Green,E.D.  
TITLE Human chromosome 7 STSs (1997)  
JOURNAL Unpublished (1997)  
COMMENT On Apr 3, 1997 this sequence version replaced gi:1706935.  
Synonyms: RELN  
GDB.DSEG: RELN

Contact: Eric D. Green  
Genome Technology Branch  
National Human Genome Research Institute/NIH  
49 Convent Dr., MSC4431, Bldg. 49, Rm. 2A08, Bethesda, MD 20892  
Tel: 3014020201  
Fax: 3014024735  
Email: egreen@nhgri.nih.gov  
Primer A: CTCCTACCTTCTCTGAGGACGCCAA  
Primer B: AAGCAGTTGCTGTGTCCACTG  
STS size: 183  
PCR Profile:  
Presoak: 0 degrees C for 0.00 minute(s)  
Denaturation: 92 degrees C for 0.17 minute(s)  
Annealing: 55 degrees C for 1.00 minute(s)  
Polymerization: 72 degrees C for 1.00 minute(s)  
PCR Cycles: 35  
Thermal Cycler: PerkinElmer 9600  
Protocol:  
Template: 30-100 ng  
Primer: each 1 uM  
dNTPs: each 200 uM  
Taq Polymerase: 0.05 units/ul  
Total Vol: 10 ul

Buffer: MgCl2: 1.5 mM  
KCl: 100 mM  
Tris-HCl: 10 mM  
NH4Cl: 5 mM  
pH: 8.6

This STS was developed from sequence determined by another investigator. See GenBank record: U79716 For additional information about the NHGRI chromosome 7 mapping project, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>. Also see Genomics 11:548-64 (1991) [MUID=92128937].

FEATURES  
Location/Qualifiers  
1..11580  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="7"  
/clone\_lib="Eric D. Green"  
1..11580  
/gene="RELN"  
1112..1294  
/gene="RELN"  
1112..1134  
/gene="RELN"  
primer\_bind complement(1272..1294)  
primer\_bind 3014 a 2696 c 2753 g 3116 t  
BASE COUNT  
ORIGIN  
Alignment Scores:



Pred. No.: 1.12e-18 Length: 11580  
 Score: 27.00 Matches: 27  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 23.08% Indels: 0  
 DB: 11 Gaps: 0

US-09-897-438B-2 (1-117) x G30936 (1-11580)

Qy 28 LeuAsnThrThrAlaSerValLeuGlnPheSerIleGlySerGlySerCysArgPhe 47  
 |||||  
 Db 941 CTTAATACACACAGCTCTGTCTCCAAATTTCCATTTGGTCAGGTTTCATGTCGCTTT 1000

Qy 48 SerTyrSerAspProSerIle 54  
 |||||  
 Db 1001 AGTTATTACAGCCCGCAGCATC 1021

RESULT 9  
 G30938 11580 bp DNA linear STS 28-SEP-1998  
 LOCUS SWSS3176 Eric D. Green Homo sapiens STS genomic, sequence tagged  
 DEFINITION

ACCESSION G30938.1 GI:1923211  
 VERSION G30938  
 KEYWORDS STS.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS  
 TITLE A collection of 1814 human chromosome 7-specific STSs  
 JOURNAL Genome Res. 7 (1), 59-64 (1997)  
 MEDLINE 97189344  
 PUBMED 9037602

REFERENCE  
 AUTHORS  
 TITLE Human chromosome 7 STSs (1997)  
 JOURNAL Unpublished (1997)  
 COMMENT On Apr 3, 1997 this sequence version replaced gi:1706937.  
 Synonyms: REIN  
 GDB\_SEQ: REIN  
 Contact: Eric D. Green  
 Genome Technology Branch  
 National Human Genome Research Institute/NIH  
 49 Convent Dr., MSC4431, Bldg. 49, Rm. 2A08, Bethesda, MD 20892  
 Tel: 3014020201  
 Fax: 3014024735  
 Email: egreen@nhgri.nih.gov  
 Primer A: TGTGGCTGATTACTAACAG  
 Primer B: TGGGCTTTTATAGCTTG  
 STS size: 238  
 PCR Profile:

Presoak: 0 degrees C for 0.00 minute(s)  
 Denaturation: 92 degrees C for 0.17 minute(s)  
 Annealing: 50 degrees C for 1.00 minute(s)  
 Polymerization: 72 degrees C for 1.00 minute(s)  
 PCR Cycles: 35  
 Thermal Cycler: PerkinElmer 9600

Protocol:  
 Template: 30-100 ng  
 Primer: each 1 uM  
 dNTPs: each 200 uM  
 Taq Polymerase: 0.05 units/ul  
 Total Vol: 10 ul

Buffer:  
 MgCl2: 1.5 mM  
 KCl: 100 mM  
 Tris-HCl: 10 mM  
 NH4Cl: 5 mM

pH: 8.6

This STS was developed from sequence determined by another investigator. See GenBank record: U79716 For additional information about the NHGRI chromosome 7 mapping project, see <http://www.nhgri.nih.gov/DIR/OTB/CHR7>. Also see Genomics 11:548-64 (1991) [MUID=92128937].

FEATURES  
 Location/Qualifiers  
 source  
 1..11580  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="7"  
 /clone\_lib="Eric D. Green"  
 1..11580  
 /gene="REIN"  
 11314..11551  
 /gene="REIN"  
 11314..11333  
 /gene="REIN"  
 primer\_bind  
 3014 a 2696 c 2753 g 3116 t 1 others  
 BASE COUNT  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1.12e-18 Length: 11580  
 Score: 27.00 Matches: 27  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 23.08% Indels: 0  
 DB: 11 Gaps: 0

US-09-897-438B-2 (1-117) x G30938 (1-11580)

Qy 28 LeuAsnThrThrAlaSerValLeuGlnPheSerIleGlySerGlySerCysArgPhe 47  
 |||||  
 Db 941 CTTAATACACACAGCTCTGTCTCCAAATTTCCATTTGGTCAGGTTTCATGTCGCTTT 1000

Qy 48 SerTyrSerAspProSerIle 54  
 |||||  
 Db 1001 AGTTATTACAGCCCGCAGCATC 1021

RESULT 10  
 AC124933/c  
 LOCUS AC124933 183641 bp DNA linear HTG 31-JUL-2002  
 DEFINITION Rattus norvegicus clone CH230-466N17, \*\*\* SEQUENCING IN PROGRESS  
 ACCESSION AC124933  
 VERSION AC124933.2 GI:21952716  
 KEYWORDS HTG; HTGS\_PHASE1.  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE  
 AUTHORS  
 1 (bases 1 to 183641)  
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Ayele,M., Banks,T., Barbia,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleaveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Devila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Depn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hoques,Y., Holloway,C., Hollins,B., Homs1,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,

[illegible]

* 63316	63415: gap of unknown length
* 63416	64571: contig of 1156 bp in length
* 64572	64671: gap of unknown length
* 64672	66083: contig of 1412 bp in length
* 66084	66183: gap of unknown length
* 66184	67516: contig of 1333 bp in length
* 67517	67616: gap of unknown length
* 67617	69820: contig of 2204 bp in length
* 69821	69920: gap of unknown length
* 69921	71261: contig of 1341 bp in length
* 71262	71361: gap of unknown length
* 71362	72775: contig of 1414 bp in length
* 72776	72875: gap of unknown length
* 72876	73298: contig of 2423 bp in length
* 73299	75398: gap of unknown length
* 75399	77094: contig of 1696 bp in length
* 77095	77194: gap of unknown length
* 77195	78834: contig of 1640 bp in length
* 78835	78935: gap of unknown length
* 78935	79889: contig of 1055 bp in length
* 79930	80089: gap of unknown length
* 80090	81774: contig of 1685 bp in length
* 81775	81874: gap of unknown length
* 81875	83255: contig of 1381 bp in length

Alignment Scores:

Pred. No.:	2,66e-13	Length:	183641
Score:	23.00	Matches:	23
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	19.66%	Indels:	0
DB:	2	Gaps:	0

US-09-897-438b-2 (1-117) x AC124933 (1-183641)

Qy 1 GluInCysGlyThrIleMetHisGlyAsnAlaValThrPheCysGluProTyrGlyPro 20  
|||||  
Db 92916 GNACACTGTGGCACCATCATGCATGGCAACGCTGTCACCTTCTGTGACCATATAGTGCCT 92857  
Qy 21 ArgGluLeu 23  
|||||  
Db 92856 CGAGAGTTG 92848

RESULT 11  
AC095877  
DEFINITION Rattus norvegicus clone CH230-10G13, \*\*\* SEQUENCING IN PROGRESS HTG 11-JUL-2002  
\*\*\*, 59 unordered pieces.

AC095877

VERSION AC095877.4 GI:21722955

KEYWORDS HTG: HTGS\_PHASE1.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 126130)

Muzy,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayele,M., Banks,T.,  
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,  
Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,  
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,  
Homsli,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,  
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
Karlssoon,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,  
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,  
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,  
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,  
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,  
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,  
Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,  
Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,  
Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H.,  
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,  
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,  
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,  
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
Weinstock,G. and Gibbs,R.

Direct Submission  
Unpublished  
2 (bases 1 to 126130)  
Worley,K.C.  
Direct Submission  
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 126130)  
Worley,K.C.  
Direct Submission  
Submitted (11-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 10, 2002 this sequence version replaced gi:20975941.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- project Information  
Center project name: GDRS  
Center clone name: CH230-10G13  
----- Summary Statistics  
Sequencing vector: Plasmid:  
Chemistry: Dye-terminator Big Dye 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 70934 bases at least Q40  
Consensus quality: 75985 bases at least Q30  
Consensus quality: 80731 bases at least Q20  
-----

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
consists of 59 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 1342: contig of 1382 bp in length  
\* 1383  
\* 1483  
\* 3046: contig of 1564 bp in length  
\* 3047  
\* 3146: gap of unknown length  
\* 3147  
\* 4489: contig of 1343 bp in length  
\* 4490  
\* 4590: gap of unknown length  
\* 4590  
\* 5786: contig of 1197 bp in length  
\* 5787  
\* 5886: gap of unknown length

```
* 5887 7115: contig of 1229 bp in length
* 7116 7215: gap of unknown length
* 7216 8771: contig of 1556 bp in length
* 8772 8871: gap of unknown length
* 8873 10260: contig of 1389 bp in length
* 10261 10360: gap of unknown length
* 10361 11370: contig of 1010 bp in length
* 11371 11470: gap of unknown length
* 11471 12710: contig of 1240 bp in length
* 12711 12810: gap of unknown length
* 12811 14180: contig of 1370 bp in length
* 14181 14280: gap of unknown length
* 14281 16048: contig of 1768 bp in length
* 16049 16148: gap of unknown length
* 16149 17569: contig of 1421 bp in length
* 17570 17669: gap of unknown length
* 17670 18820: contig of 1151 bp in length
* 18821 18920: gap of unknown length
* 18921 20209: contig of 1289 bp in length
* 20210 20309: gap of unknown length
* 20310 21633: contig of 1324 bp in length
* 21634 21733: gap of unknown length
* 21734 23269: contig of 1536 bp in length
* 23270 23369: gap of unknown length
* 23370 24647: contig of 1278 bp in length
* 24648 24747: gap of unknown length
* 24748 25800: contig of 1053 bp in length
* 25801 25900: gap of unknown length
* 25901 27017: contig of 1117 bp in length
* 27018 27117: gap of unknown length
* 27118 29183: contig of 2066 bp in length
* 29184 29283: gap of unknown length
* 29284 30315: contig of 1032 bp in length
* 30316 30415: gap of unknown length
* 30416 32086: contig of 1671 bp in length
* 32087 32186: gap of unknown length
* 32187 33395: contig of 1209 bp in length
* 33396 33495: gap of unknown length
* 33496 35663: contig of 2168 bp in length
* 35664 35763: gap of unknown length
* 35764 38044: contig of 2281 bp in length
* 38045 38144: gap of unknown length
* 38145 39845: contig of 1701 bp in length
* 39846 39945: gap of unknown length
* 39946 41982: contig of 2037 bp in length
* 41983 42083: gap of unknown length
* 42083 43743: contig of 1661 bp in length
* 43744 43843: gap of unknown length
* 43844 45904: contig of 2061 bp in length
* 45905 46004: gap of unknown length
* 46005 47667: contig of 1663 bp in length
* 47668 47767: gap of unknown length
* 47768 49833: contig of 2066 bp in length
* 49834 49933: gap of unknown length
* 49934 51434: contig of 1501 bp in length
* 51435 53778: gap of unknown length
* 53779 53778: contig of 2244 bp in length
* 53779 53778: gap of unknown length
* 53779 55552: contig of 1674 bp in length
* 55553 55652: gap of unknown length
* 55653 57670: contig of 2018 bp in length
* 57671 57770: gap of unknown length
* 57771 59630: contig of 1860 bp in length
* 59631 59730: gap of unknown length
* 59731 61147: contig of 1417 bp in length
* 61148 61247: gap of unknown length
* 61248 63995: contig of 2748 bp in length
* 63996 64096: gap of unknown length
* 64096 66294: contig of 2199 bp in length
* 66295 66394: gap of unknown length
* 66395 68894: contig of 2400 bp in length
* 68895 71533: contig of 2639 bp in length

* 71534 71634: gap of unknown length
* 71634 73470: contig of 1836 bp in length
* 73470 73569: gap of unknown length
* 73569 77054: contig of 3485 bp in length
* 77054 77154: gap of unknown length
* 77155 79005: contig of 1851 bp in length
* 79006 79105: gap of unknown length
* 79106 81295: contig of 2189 bp in length
* 81295 81394: gap of unknown length
* 81394 83964: contig of 2470 bp in length
* 83965 83964: gap of unknown length
* 83965 86039: contig of 2075 bp in length
* 86040 86139: gap of unknown length
* 86140 88646: contig of 2507 bp in length
* 88647 88746: gap of unknown length
* 88747 91823: contig of 3077 bp in length
* 91824 91923: gap of unknown length
* 91924 94369: contig of 2446 bp in length
* 94370 94469: gap of unknown length
* 94470 96886: contig of 2417 bp in length
* 96887 96986: gap of unknown length
* 96987 100679: contig of 3693 bp in length
* 100680 100779: gap of unknown length
* 100780 104171: contig of 3392 bp in length
* 104172 104271: gap of unknown length
* 104272 106760: contig of 2489 bp in length

Alignment Scores:
Pred. No.: 2,3e-12 Length: 126130
Score: 22.00 Matches: 22
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 18.80% Indels: 0
DB: 2 Gaps: 0

US-09-897-438B-2 (1-117) x AC095877 (1-126130)
Qy 96 ValGlnPheGlnTrpLysGlnAspSerLeuArgValGlyGluValTyrGluAlaCysTrp 115
|||||
Db 92896 GTGCAGTTCAGTGGAAACAGGACGCTGCTGTGGTGGTGTACGAGCCTGCTGG 92955

Qy 116 AlaLeu 117
|||||
Db 92956 GCCCTG 92961

RESULT 12
AC128022/c 202764 bp DNA linear HTG 19-JUL-2002
LOCUS Rattus norvegicus clone CH230-525L20, *** SEQUENCING IN PROGRESS
DEFINITION *** 97 unordered pieces.
ACCESSION AC128022
VERSION AC128022.1 GI:21908605
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 202764)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davilla,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J.J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
```

Gorrell, J. H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,  
 Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,  
 Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L. E.,  
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,  
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L.,  
 Li, J., Li, Z., Lichtarge, O., Lied, C., Liu, J., Liu, W., Loulseged, H.,  
 Lozadow, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,  
 Massey, E., Mawhiney, E., McLeod, M. P., Meador, M., Mei, G., Metzker, M.,  
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G.,  
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
 Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y.,  
 Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G.,  
 Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I.,  
 Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,  
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
 Tansley, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
 Usmani, K., Vasquez, L., Vera, V., Villaion, D., Vinson, R., Wang, Q.,  
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
 Williams, G., Williamson, A., Wleczyk, R., Woodson, S., Worley, K.,  
 Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstein, G., and Gibbs, R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 202764)  
 Worley, K. C.  
 Direct Submission  
 Submitted (19-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
 ----- Project Information  
 Center project name: KAI1  
 Center clone name: CH230-525L20  
 ----- Summary Statistics  
 Sequencing vector: Plasmid  
 Chemistry: Dye-terminator Big Dye: 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 113069 bases at least Q40  
 Consensus quality: 119970 bases at least Q30  
 Consensus quality: 125656 bases at least Q20  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 97 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1  
 \* 1244: contig of 1244 bp in length  
 \* 1245  
 \* 1344: gap of unknown length  
 \* 1345  
 \* 2346: contig of 1002 bp in length  
 \* 2347  
 \* 2446: gap of unknown length  
 \* 2447  
 \* 3813: contig of 1367 bp in length  
 \* 3814  
 \* 3913: gap of unknown length  
 \* 3914  
 \* 5101: contig of 1188 bp in length  
 \* 5102  
 \* 5201: gap of unknown length  
 \* 5202  
 \* 6724: contig of 1523 bp in length  
 \* 6725  
 \* 6825: gap of unknown length  
 \* 6825  
 \* 8048: contig of 1224 bp in length  
 \* 8049  
 \* 8148: gap of unknown length  
 \* 8149  
 \* 9212: contig of 1064 bp in length

9312: gap of unknown length  
 9313  
 10432: contig of 1120 bp in length  
 10433  
 10532: gap of unknown length  
 10533  
 11890: contig of 1358 bp in length  
 11891  
 11990: gap of unknown length  
 11991  
 13058: contig of 1078 bp in length  
 13059  
 13168: gap of unknown length  
 13169  
 14486: contig of 1318 bp in length  
 14487  
 14586: gap of unknown length  
 14587  
 15996: contig of 1410 bp in length  
 15997  
 16096: gap of unknown length  
 16097  
 17534: contig of 1438 bp in length  
 17535  
 17634: gap of unknown length  
 17635  
 18703: contig of 1069 bp in length  
 18704  
 18803: gap of unknown length  
 18804  
 19884: contig of 1081 bp in length  
 19885  
 19984: gap of unknown length  
 19985  
 21539: contig of 1555 bp in length  
 21540  
 21639: gap of unknown length  
 21640  
 22780: contig of 1141 bp in length  
 22781  
 22880: gap of unknown length  
 22881  
 23926: contig of 1046 bp in length  
 23927  
 24026: gap of unknown length  
 24027  
 25444: contig of 1518 bp in length  
 25445  
 25644: gap of unknown length  
 25645  
 26899: contig of 1255 bp in length  
 26900  
 26999: gap of unknown length  
 27000  
 28237: contig of 1238 bp in length  
 28238  
 28337: gap of unknown length  
 28338  
 29758: contig of 1421 bp in length  
 29759  
 31038: contig of 1180 bp in length  
 31039  
 31138: gap of unknown length  
 31139  
 32179: contig of 1041 bp in length  
 32180  
 32279: gap of unknown length  
 32280  
 33796: contig of 1517 bp in length  
 33797  
 33896: gap of unknown length  
 33897  
 35380: contig of 1484 bp in length  
 35381  
 35480: gap of unknown length  
 35481  
 37032: contig of 1552 bp in length  
 37033  
 37132: gap of unknown length  
 37133  
 38414: contig of 1282 bp in length  
 38415  
 38514: gap of unknown length  
 38515  
 39948: contig of 1434 bp in length  
 39949  
 40048: gap of unknown length  
 40049  
 41436: contig of 1388 bp in length  
 41437  
 41536: gap of unknown length  
 41537  
 43084: contig of 1548 bp in length  
 43085  
 43184: gap of unknown length  
 43185  
 44940: contig of 1756 bp in length  
 44941  
 45040: gap of unknown length  
 45041  
 46186: contig of 1146 bp in length  
 46187  
 46286: gap of unknown length  
 46287  
 47989: contig of 1703 bp in length  
 47990  
 48089: gap of unknown length  
 48090  
 49842: contig of 1753 bp in length  
 49843  
 49942: gap of unknown length  
 49943  
 51128: contig of 1186 bp in length  
 51129  
 51228: gap of unknown length  
 51229  
 52823: contig of 1595 bp in length  
 52824  
 52923: gap of unknown length  
 52924  
 54431: contig of 1508 bp in length  
 54432  
 54531: gap of unknown length  
 54532  
 56042: contig of 1511 bp in length  
 56043  
 56142: gap of unknown length  
 56143  
 57287: contig of 1145 bp in length  
 57288  
 57387: gap of unknown length  
 57388  
 58893: contig of 1506 bp in length  
 58894  
 58993: gap of unknown length  
 58994  
 60024: contig of 1031 bp in length  
 60025  
 60124: gap of unknown length  
 60125  
 62858: contig of 2734 bp in length  
 62859

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

* 62959	64365:	contig of 1407 bp in length
* 64366	64465:	gap of unknown length
* 64466	64434:	contig of 1969 bp in length
* 66435	66534:	gap of unknown length
* 66435	66534:	gap of unknown length
* 66535	66631:	contig of 2097 bp in length
* 68632	68731:	gap of unknown length
* 68732	70939:	contig of 2108 bp in length
* 70840	70939:	gap of unknown length
* 70940	73109:	contig of 2070 bp in length
* 73109	73109:	gap of unknown length
* 73110	74957:	contig of 1848 bp in length
* 74958	75057:	gap of unknown length
* 75058	76816:	contig of 1759 bp in length
* 76817	76916:	gap of unknown length
* 76917	76979:	contig of 1763 bp in length
* 78680	78779:	gap of unknown length
* 78780	80056:	contig of 1277 bp in length
* 80057	80156:	gap of unknown length
* 80157	81312:	contig of 1156 bp in length
* 81313	81412:	gap of unknown length
* 81413	82922:	contig of 1510 bp in length
* 82923	83022:	gap of unknown length
* 83023	84180:	contig of 1158 bp in length
* 84181	84280:	gap of unknown length
* 84281	85849:	contig of 1569 bp in length
* 85850	85949:	gap of unknown length
* 85950	87612:	contig of 1663 bp in length
* 87613	87712:	gap of unknown length

Alignment Scores:		
Pred. No.:	3,49e-12	Length: 202764
Score:	22.00	Matches: 22
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	18.80%	Indels: 0
DB:	2	Gaps: 0

US-09-897-438b-2 (1-117) x AC128022 (1-202764)

Qy 96 ValGlnPheGlnTnTlPlySAspSerLeuArgValGlyGluValTyGluAlaCysTrp 115  
|||||  
Db 187246 GTGCAGTTCGAGTCCGAAACAGGACGCTCGTGTGGTGGAGGTACGAGCCCTGCTGG 187187

Qy 116 AlaLeu 117

|||||

Db 187186 GCCCTG 187181

RESULT 13  
AF090441  
LOCUS AF090441 10634 bp mRNA linear VRT 15-SEP-1998  
DEFINITION Gallus gallus extracellular reelin (Rein) mRNA, partial cds.  
ACCESSION AF090441  
VERSION AF090441.1 GI:3600101

KEYWORDS Gallus gallus.

SOURCE Gallus gallus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 10634)

AUTHORS Bernier,B. and Goffinet,A.M.

TITLE Comparative study of reelin in vertebrates

JOURNAL Unpublished

AUTHORS Bernier,B. and Goffinet,A.M.

TITLE Direct Submission

JOURNAL Submitted (04-SEP-1998) Neurobiology, Fac.N-D de la Paix, Rue de Bruxelles, 61, Namur B-5000, Belgium

FEATURES Location/Qualifiers

source 1..10634

/organism="Gallus gallus"

/db\_xref="taxon:9031"

gene <1..10634

CDs  
  
/gene="Rein"  
<1..9631  
/gene="Rein"  
/codon\_start=2  
/product="extracellular reelin"  
/protein\_id="AAC35559.1"  
/db\_xref="GI:3600102"  
/translation="TGLNTTTASVLQFSLGSGCRFSYSDPSITVSYSKNSADWTQ  
LEKISAPSNVTIILHLYLPEDAKGENVHFQWKODYLHAGEVYEACWALDNLILINAA  
HRKVLEDNLDVDVTCGNLWLFPPCATVKHSCQDGNISYPHGTGSEFENATITPMDVLS  
TEDAQAEFEESQPKGWDILGAVIGTECQLESGSSWFLDGRKKICTPTMDVTIG  
YGNLRFYSGMGNCDSGSHENDVILYAKIEGRREHIALDTITLAAKYLSVSVVVIS  
PDLQTPATKFCCLKQKSHQGHNRNVAWDFHVLPLVPTTHMIOFISINLGGCTVOPG  
NSVSEFSTNHRGWSLLHTECLPEICAGPHLPHSTVYASENYSGNRTTTPVPAAL  
TSDTRIMRQTGPIHGNMNAIDNIYIGPSCLKFCSCRGCTKNGCKDCDGFSGACET  
ASQTFPMFISEFASRLSSYHNFIYIRGAEVSGCGVLASGKALVFNKDKRQLITA  
FLDSSQREFLQFTLRUGSKSVJSTKAPDQPGEGVILLHYSYDNGITWKLHYSYLN  
HEPRIISVELPDAQIGQFRWQVPHSSQSGEDWAIIDEIVMTSYLNFNSISLDTNL  
VEVTSLGYLGNVQPYCGHDMTLCFTGDSKLTSSMYVETSMQISAGYMQFNLM  
GCGQKFTPHMDNOVKLEYSTNHGLTWLHVOECLPSMPCOEFTSASIVHSEFTQWR  
RITVLLPQKTSASATFRWSQCYTAPDEWALDNIYIGQCPNMGSGHMGHGVCR  
DSGFRCTEQPENPLSTVMSDEENPDVLKTEQELIGGEIVKPEGCGVYISGSSLY  
FNKAGRQLYSMDLTTWVDFVOFYIQIGSESSCNRPSREBVLQIYNSNGINWQ  
LLAEMVYDFSKPRFYLELPAAAKTCTFRFRWQPVFSGEDQWAIIDITLSKQ  
KHIIIPVNPFLQNFYKPAFYPMNLQSVWLLILANEGMTKNESFCFSATMLFKS  
DGRFAVTRDLTKPGYVLQFKLINIGCTNOYSSAPVLLQYSHDAGLFWSLKGCYP  
ASPTKGCSSSRELSEPTVYHTGDFEDWTRITIVIPRSLAAKTRFRWIOESSHKS  
VPPGLDGYIISEPCNYONGHCDGVSGVCFCDLGYTASHGTCVSNVPMHSMFDFE  
RKLSPLYAKITGCGVGTGCVLSDGASLYFNGPGKREARTVPLDTNIRLVQFYVQIG  
SKATGSCNRPSRNEGLIVQYNDNGITWHLRLDIFMSYLEPQVVSIDLPREAKTS  
ATAFRWQPHGHSQAQWALDDVLIGMDSQSQTGFQDFGVDTDLQASRYITGGQVD  
IDCLMDTALMFSENIKEPRYAETWDFHVSASTFLFELSMGCSKPSYSHSTHLYS  
LNGRDHLVTECPVPTIGCOHYTESSTYTERFONKRIITAYILPITNSPTRPRW  
IOYVASGVDSWAINVLAATGCPNMGSGHCDGACVCDRGGFGPYCVHNPPLSV  
LKDENGNLHPLDWPYGAERGNLNGDTIKSGTALLFKGEGRLMLVSRDLCTNTVY  
IQSFPAKGTPEPERSHILQISVNGGITWHLIDFETITDQVLFILNPLPYTAQS  
NATFRRLWQPYNSGKEEIIIDDFIIDGNLKNPILIDTDFGPKEDNWFYPOGN  
IGLYCPSYSGKGAPEEDSAMVFSYNEVSEHSITRDLNVENITIQIEINIGCTDSSS  
ADPKLEFSRDGATWHLILPLYCSSSHSLSLSCSTEHPHSTTYTGTGWRREVIH  
FKGLHLCTFRWYOGFYAGSOPVTAIDNVYIGPOCEEMCGHSGCINGTKCICD  
PGYSGPTCKISTKNSDSLKDDPEGQLESDFRLVYSGGKPSKCGKMGSGGNLFFBEG  
LNMTRDLQDLQARFVQFMRLGCGKVPDRSQLSLQYSLNGGLTWSLQGEFLFS  
NNSNGRYIALEIPMKARSSSTRLWQVSENGHFSYPMWIDQILIGGISSTVLED  
ASCVTGSCYAIETELYSVDLITWHPILRDLPTNYECNRYHLRLISDTFNKWTRI  
TLPLPYTSQATFRFRWOPADPKOOWAIDNVYIGDGCIDMCSGHGKCTODNCVCD  
EHWGLYCDPEPETPLTQLKDNFNSPQNMLTVNGKLSVCGAVASGMAHFESEG  
CSRMLVTDNLITNAEFTQFYMYGCLITPNRNQGVLLIYSVNGGITSPLMEIFYD  
QFSKPGFVNILLIPYDAKTIGTRFRWQPKHGLDQNDWAIIDNVLSIDQRTVMDLT  
FSAPLPQHRSPPADAGPTGRIAFDMFMEDKTVNEHWFHDDCSTERCDSPDGMYI  
CGSHDREYVAVTHTLPTTEGIMQFKVSGCKTSEKLAQNQVHVQYSTDFGWSYI  
VPQCLPADPKCSGSOPSVPFPTKWKRVYSLPENLVGNPVRFRFYOKYSDOWAI  
DNFYLGPGCLENGRCHGDCLEKQICDPCYSGPCYLTOTLTKFLKERDNEIKPDL  
WMSLEGNTCEGCLAEEDTTLFGGQTVQAVTQDLRGAFLQYWGRISENNMT  
TCHRPTRKREGLVSDIGGITWTLHEMDYOKYISVRHXYILPEHALTNTLRUR  
WQPFYISNGIVVSGPDQAWALDNLIGGAENPSQVDTFDEGTSHEENSSYPNA  
VTAAGCPNPSFLHYPNKKDKTHNLSRELIQGYMQFKIYVVGCSGSCGLH  
SVMLYTKDARTDSWOLVOTHCLPSSNSIGCSPPQFHEATVYNSVNSMWRRTITQL  
PDHVSSATQFRWIOKGELEKQSWAIDNIYIGAGPKLCSGRGYCSTCAICIDGEG  
QGDSCVSFSDHLPSTIKDNFESERVEITENWETIQGGVINGCGQLAPYAHGSLYNG  
CQVRQAVTKHQPDKFIIQARVSYNPVLEARNKGVLLRWMQPRHNGTGHQWALDHVEVVL  
TQWIAHQHPKDFIIQARVSYNPVLEARNKGVLLRWMQPRHNGTGHQWALDHVEVVL  
ISTRKQNYMMNFSRQHLRHFYNNRRRLRRPY"

BASE COUNT	2839 a	2392 c	2540 g	2862 t	1 others
ORIGIN					
Alignment Scores:					
Pred. No.:	5,48e-09	Length: 10634			
Score:	18.00	Matches: 18			
Percent Similarity:	100.00%	Conservative: 0			
Best Local Similarity:	100.00%	Mismatches: 0			
Query Match:	15.38%	Indels: 0			
DB:	5	Gaps: 0			



```

repeat_region 40278. .40311
/rpt_family="L1"
repeat_region complement(41921. .42075)
/rpt_family="ALU"
repeat_region complement(42447. .42734)
/rpt_family="ALU"
exon 43126. .43185
/note="GRAIL prediction, score = 95"
/evidence=not_experimental
repeat_region complement(43411. .43680)
/rpt_family="ALU"
repeat_region 45704. .45991
/rpt_family="ALU"
misc_feature 46064. .46434
/note="match to Human cDNA H60097, 5' end (NID:g1012929)"
repeat_region complement(46604. .46897)
/rpt_family="ALU"
misc_feature complement(47023. .47412)
/gene="WUGSC:H_RG249A12.1"
/note="match to Human cDNA, 3' end H59444 (NID:g1012276)"
repeat_region 58125. .58163
/rpt_family="L1"
repeat_region 58604. .58758
/rpt_family="L1"
repeat_region 59743. .59765
/rpt_family="L1"
repeat_region 60210. .60516
/rpt_family="ALU"
exon complement(63051. .63146)
/gene="WUGSC:H_RG249A12.1"
/evidence=not_experimental
repeat_region 64242. .64284
/rpt_family="L1"
repeat_region 64445. .64496
/rpt_family="L1"
repeat_region complement(67753. .68011)
/rpt_family="L1"
repeat_region 67986. .68889
/rpt_family="L1"
exon complement(70725. .70841)
/gene="WUGSC:H_RG249A12.1"
/evidence=not_experimental
repeat_region 71911. .71947
/rpt_family="L1"
repeat_region complement(73631. .73919)
/rpt_family="ALU"
repeat_region 75073. .75092
/rpt_family="L1"
repeat_region 76743. .82366
/rpt_family="L1"
exon 78326. .79402
/note="GRAIL prediction, score = 83"
/evidence=not_experimental
exon 80386. .81548
/note="GRAIL prediction, score = 82"
/evidence=not_experimental
repeat_region complement(80962. .81380)
/rpt_family="L1"
repeat_region complement(82581. .82608)
/rpt_family="L1"
misc_feature 83877. .84447

```

## Alignment Scores:

```

Pred. No.: 3 69e-08 Length: 93163
Score: 18.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 15.38% Indels: 0
DB: 9 Gaps: 0

```

US-09-897-438B-2 (1-117) x HSAC000121 (1-93163)

```

Oy 6 IleMethHisGlyAsnAlaValThrPheCysGluProTyGlyProArgGluLeu 23
|||||
Db 63103 ATTATGCATGCGTCGCTCACCTTCTGTGAACCATATGCCCCAGAGACTG 63050

RESULT 15
AC041023/c
LOCUS AC041023 Homo sapiens clone RP11-356G3, WORKING DRAFT SEQUENCE, 20 unordered
DEFINITION pieces.
AC041023
VERSION AC041023.2 GI:8016758
HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
1 (bases 1 to 185996)
Homo sapiens, clone RP11-356G3
Unpublished
2 (bases 1 to 185996)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bedar,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burckett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (11-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 22, 2000 this sequence version replaced gi:7534225.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIGR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9601
Center clone name: 356.G.3
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 175155 bases at least Q40
Consensus quality: 180623 bases at least Q30
Consensus quality: 182712 bases at least Q20
Insert size: 188000; agarose-fp
Quality coverage: 4.5 in Q20 bases; agarose-fp
Quality coverage: 4.6 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is

```



\* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

1      2941: contig of 2941 bp in length
*      2942 3041: gap of 100 bp
*      3042 5193: contig of 2152 bp in length
*      5194 5293: gap of 100 bp
*      5294 7050: contig of 1757 bp in length
*      7051 7150: gap of 100 bp
*      7151 10452: contig of 3302 bp in length
*      10453 10552: gap of 100 bp
*      10553 13859: contig of 3307 bp in length
*      13860 13959: gap of 100 bp
*      13960 17254: contig of 3295 bp in length
*      17255 17354: gap of 100 bp
*      17355 20523: contig of 3169 bp in length
*      20524 20623: gap of 100 bp
*      20624 25148: contig of 4525 bp in length
*      25149 25248: gap of 100 bp
*      25249 29676: contig of 4428 bp in length
*      29677 29776: gap of 100 bp
*      29777 34077: contig of 4301 bp in length
*      34078 34177: gap of 100 bp
*      34178 39009: contig of 4832 bp in length
*      39010 39109: gap of 100 bp
*      39110 45564: contig of 6455 bp in length
*      45565 45664: gap of 100 bp
*      45665 54035: contig of 8371 bp in length
*      54036 54135: gap of 100 bp
*      54136 63394: contig of 9259 bp in length
*      63395 63494: gap of 100 bp
*      63495 74946: contig of 11452 bp in length
*      74947 75046: gap of 100 bp
*      75047 88353: contig of 13307 bp in length
*      88354 88453: gap of 100 bp
*      88454 104853: contig of 16400 bp in length
*      104854 104953: gap of 100 bp
*      104954 123339: contig of 18386 bp in length
*      123340 123439: gap of 100 bp
*      123440 154730: contig of 31291 bp in length
*      154731 154830: gap of 100 bp
*      154831 185996: contig of 31166 bp in length.

```

## FEATURES

## Source

```

1. .185996
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="RP11-356G3"
/clone_lib="RPC1-11 Human Male BAC"
misc_feature
1. .2941
/note="assembly_fragment"
misc_feature
3042. .5193
/note="assembly_fragment"
misc_feature
5294. .7050
/note="assembly_fragment"
misc_feature
7151. .10452
/note="assembly_fragment"
misc_feature
10553. .13859
/note="assembly_fragment"
misc_feature
13960. .17254
/note="assembly_fragment"
misc_feature
17355. .20523
/note="assembly_fragment"
misc_feature
20624. .25148
/note="assembly_fragment"
misc_feature
25249. .29676
/note="assembly_fragment"
misc_feature
29777. .34077
/note="assembly_fragment"
misc_feature
34178. .39009

```

```

misc_feature
/note="assembly_fragment"
39110. .45564
/note="assembly_fragment"
45665. .54035
/note="assembly_fragment"
54136. .63394
/note="assembly_fragment"
63495. .74946
/note="assembly_fragment"
75047. .88353
/note="assembly_fragment"
88454. .104853
/note="assembly_fragment"
104954. .123339
/note="assembly_fragment"
123440. .154730
/note="assembly_fragment"
154831. .185996
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
BASE COUNT 60145 a 34255 c 33646 g 56049 t 1901 others
ORIGIN

```

```

Alignment Scores:
Pred. No.: 6.78e-08 Length: 185996
Score: 18.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 15.38% Indels: 0
DB: 2 Gaps: 0

```

US-09-897-438B-2 (1-117) x AC041023 (1-185996)

Qy 6 IleMethHisGlyAsnAlaValThrPheCysGluProTyrGlyProArgGluLeu 23

```

|||||
Db 166357 ATTATGCATGGCATGGCGTCACCTTCTGTGACCATATGCCCCCAGAGAACTG 166304

```

Search completed: November 6, 2002, 23:00:38  
 Job time : 2904 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 16:42:01 ; Search time 68 Seconds  
(without alignments)  
1582.993 Million cell updates/sec

Title: US-09-897-438B-1  
Perfect score: 351  
Sequence: 1 gagcagtggtgacccatcat.....atgagcctgtggccctg 351

Scoring table: OLIGO.NUC  
Gap 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues  
Word size: 0  
Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: Issued\_Patents\_NA:  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Score	Match	Length	DB	ID	Description
1	351	100.0	11673	4	US-09-334-220-3		Sequence 3, Appli
2	26	7.4	11580	4	US-09-334-220-4		Sequence 4, Appli
3	18	5.1	1074	2	US-08-627-151A-15		Sequence 15, Appli
4	18	5.1	1404	6	5171840-8		Patent No. 5171840
5	18	5.1	1404	6	5480796-8		Patent No. 5480796
6	18	5.1	1486	4	US-08-795-473B-3		Sequence 3, Appli
7	18	5.1	1486	4	US-08-439-856-3		Sequence 3, Appli
8	18	5.1	2061	6	5171840-1		Patent No. 5171840
9	18	5.1	2061	6	5480796-1		Patent No. 5480796
10	18	5.1	3319	4	US-08-795-473B-2		Sequence 2, Appli
11	18	5.1	3319	4	US-08-439-856-2		Sequence 2, Appli
12	17	4.8	47	2	US-08-642-684-13		Sequence 13, Appli
13	17	4.8	1434	4	US-09-480-921B-9		Sequence 9, Appli
14	17	4.8	1434	4	US-09-480-921B-28		Sequence 28, Appli
15	17	4.8	1488	4	US-09-108-010B-11		Sequence 11, Appli
16	17	4.8	1732	4	US-09-449-335-1		Sequence 1, Appli
17	17	4.8	1732	4	US-09-449-335-5		Sequence 5, Appli
18	17	4.8	1770	2	US-08-642-684-3		Sequence 3, Appli
19	17	4.8	2935	4	US-09-480-921B-27		Sequence 27, Appli
20	17	4.8	6386	2	US-08-483-376-1		Sequence 1, Appli
21	17	4.8	8050	4	US-09-491-362-11		Sequence 11, Appli
22	17	4.8	8050	4	US-09-874-562-11		Sequence 11, Appli
23	16	4.6	2692	4	US-08-981-392-11		Sequence 11, Appli
24	16	4.6	2857	4	US-08-981-392-4		Sequence 4, Appli
25	16	4.6	3600	4	US-09-657-042A-3		Sequence 3, Appli
26	16	4.6	4181	4	US-09-393-569-1		Sequence 1, Appli
27	16	4.6	5160	3	US-08-895-601-1		Sequence 1, Appli

28	16	4.6	5430	3	US-09-012-515A-11	Sequence 11, Appli
29	16	4.6	5430	3	US-08-360-144A-11	Sequence 11, Appli
30	16	4.6	5430	4	US-09-012-504A-11	Sequence 11, Appli
31	16	4.6	6911	1	US-08-311-174-4	Sequence 4, Appli
32	16	4.6	7653	4	US-08-471-112A-1	Sequence 1, Appli
33	16	4.6	7824	5	PCT-US95-06722-11	Sequence 11, Appli
34	16	4.6	8948	4	US-09-643-597-119	Sequence 119, Appli
35	15	4.3	81	4	US-09-133-321-7	Sequence 7, Appli
36	15	4.3	307	4	US-09-615-192A-165	Sequence 165, Appli
37	15	4.3	358	2	US-08-454-557C-15	Sequence 15, Appli
38	15	4.3	358	2	US-08-340-426D-15	Sequence 15, Appli
39	15	4.3	358	2	US-08-450-673C-15	Sequence 15, Appli
40	15	4.3	358	5	PCT-US95-17111A-15	Sequence 15, Appli
41	15	4.3	463	1	US-07-752-101A-67	Sequence 67, Appli
42	15	4.3	531	2	US-08-809-267-9	Sequence 9, Appli
43	15	4.3	531	5	PCT-US95-13662A-9	Sequence 9, Appli
44	15	4.3	557	4	US-09-404-879A-88	Sequence 88, Appli
45	15	4.3	594	4	US-09-404-879A-87	Sequence 87, Appli

ALIGNMENTS

RESULT 1  
US-09-334-220-3  
Sequence 3, Application US/09334220  
Patent No. 6323177  
GENERAL INFORMATION:  
APPLICANT: St. Jude's Children's Research Hospital  
APPLICANT: Curran, Thomas  
INVENTOR: D'Arcangelo, Gabriella  
TITLE OF INVENTION: INTERACTION OF REELIN WITH VERY LOW DENSITY LIPOPROTEIN (VLDL) RECEPTOR FOR SCREENING AND TREATING  
TITLE OF INVENTION: DENSITY LIPOPROTEIN (VLDL) RECEPTOR FOR SCREENING AND TREATING  
FILE REFERENCE: 2427/0F704  
CURRENT FILING DATE: 1999-06-16  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: FASTSEQ for Windows version 3.0  
SEQ ID NO: 3  
LENGTH: 11673  
TYPE: DNA  
ORGANISM: Mus musculus  
US-09-334-220-3

Query Match 100.0%; Score 351; DB 4; Length 11673;  
Best Local Similarity 100.0%; Pred. No. 1.4e-174;  
Matches 351; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	GAGCAGTGTGGCACCACATCATGCGCAATGCTGTACCTTCTGTGAGCCGTAGGGCCCT	60
Db	970	GAGCAGTGTGGCACCACATCATGCGCAATGCTGTACCTTCTGTGAGCCGTAGGGCCCT	1029
Qy	61	CGAGAGTGTGACCCACACATGCTGAAACACACACACATCTGCTCCAGTTTCCATT	120
Db	1030	CGAGAGTGTGACCCACACATGCTGAAACACACACACATCTGCTCCAGTTTCCATT	1089
Qy	121	GGGTGAGGATCATGTCGATTTAGTTACTCTGACCCAGCATCTGTGTATACCCCAAG	180
Db	1090	GGGTGAGGATCATGTCGATTTAGTTACTCTGACCCAGCATCTGTGTATACCCCAAG	1149
Qy	181	AACAATACCGTGTGATTCAGTTCAGTGGAGAAATAGAGCCCTTCCAAATGTAGGACCA	240
Db	1150	AACAATACCGTGTGATTCAGTTCAGTGGAGAAATAGAGCCCTTCCAAATGTAGGACCA	1209
Qy	241	GTCATCCACATCTGCTACCTCCCGAGGAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGG	300
Db	1210	GTCATCCACATCTGCTACCTCCCGAGGAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGG	1269
Qy	301	AAACAGGAGCAGCTCGAGTGGGTGAGTGTATGAGGCTGTGTTGGCCCTG	351
Db	1270	AAACAGGAGCAGCTCGAGTGGGTGAGTGTATGAGGCTGTGTTGGCCCTG	1320



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 15:27:06 ; Search time 300 Seconds  
(without alignments)  
2634.838 Million cell updates/sec

Title: US-09-897-438B-1  
Perfect score: 351  
Sequence: 1 gaggcgtgtggcaccatcat.....atgaggcctgtggccttg 351

Scoring table:  
OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

W size : 0  
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	351	100.0	351	ABL40165	Mouse reelin prote
2	351	100.0	2745	AAD22754	Mus musculus trunc
3	351	100.0	11673	ABA92603	Mouse reelin encod
4	351	100.0	11673	AB199268	Mouse ischaemic co
5	26	7.4	11580	ABN96939	Gene #3437 used to
6	26	7.4	11580	ABN92604	Human reelin encod
7	26	7.4	11632	AA589484	DNA encoding novel
8	22	6.3	22	ABL40169	Mouse reelin prote
9	21	6.0	21	ABL40173	Mouse reelin prote

10	20	5.7	20	24	AAD22777	Mouse truncated re
11	19	5.4	4883	22	AAH19492	Human coding seque
12	19	5.4	7215	22	AAH19497	Human coding seque
13	19	5.4	7215	22	AAH19498	Human coding seque
14	18	5.1	18	24	ABL40174	Mouse reelin prote
15	18	5.1	975	21	AAA70701	Human interleukin
16	18	5.1	1035	21	AAZ40288	SR345 coding seque
17	18	5.1	1074	18	AAV04440	Interleukin 6 rece
18	18	5.1	1260	20	AAZ09202	Human IL-6 recepto
19	18	5.1	1486	19	AAV60296	Human interleukin-
20	18	5.1	1545	21	AAA70763	IL-6R/IL-6 fusion
21	18	5.1	1627	18	AAV97848	Human fusion polyp
22	18	5.1	2061	10	AAV90340	Sequence encoding
23	18	5.1	2066	14	AAQ41746	IL-6 receptor codi
24	18	5.1	2087	10	AAV08047	DNA contg. region
25	18	5.1	3004	21	AAZ33359	Human secreted pro
26	18	5.1	3319	17	AAV31441	Interleukin-6 rece
27	18	5.1	3319	19	AAV60295	Human interleukin-
28	18	5.1	3319	21	AAV21364	Human low adenosin
29	18	5.1	3319	21	AAV35242	Human adenosine re
30	18	5.1	3319	24	ABK84527	Human cDNA differe
31	18	5.1	3477	21	AAV09047	Fusion polypeptide
32	18	5.1	3507	21	AAV09046	Fusion polypeptide
33	18	5.1	4513	21	AAV21365	Human low adenosin
34	18	5.1	4873	21	AAV35243	Human adenosine re
35	18	5.1	9720	24	ABL56176	Hordeum vulgare va
36	18	5.1	198285	24	ABK84699	Human cDNA differe
37	18	5.1	198285	24	ABN97319	Gene #3817 used to
38	17	4.8	22	21	AAV63485	Beta-actin gene PC
39	17	4.8	47	18	AAV90826	Bacillus stearothe
40	17	4.8	325	22	AAV36445	Human musculoskele
41	17	4.8	395	18	AAV13260	N. tabacum strain
42	17	4.8	429	24	ABN95973	Gene #2471 used to
43	17	4.8	559	22	ABV63804	Human foetal liver
44	17	4.8	559	22	ABV30992	Probe #9458 for ge
45	17	4.8	559	22	AAV12314	Human brain expres

ALIGNMENTS

RESULT 1  
ABL40165  
ID ABL40165 standard; DNA; 351 BP.  
XX  
AC ABL40165;  
XX  
XX  
DT 21-MAY-2002 (first entry)  
XX  
DE Mouse reelin protein CR-50 epitope region encoding DNA SEQ ID NO:1.  
XX  
KW Mouse; reelin protein CR-50 epitope region; elucidation; neuron;  
KW cerebral disturbance; reelin protein; neuroprotective; gene; ds.  
XX  
OS Mus musculus.  
XX  
XX  
FH Key Location/Qualifiers  
FT CDS 1..351  
FT /tag= a  
FT /product= reelin protein CR-50 epitope region  
FT /note= "no start or stop codons given"

JP2002017361-A.

22-JAN-2002.

04-JUL-2000; 2000JP-0202801.

04-JUL-2000; 2000JP-0202801.

(RIKA ) RIKAGAKU KENKYUSHO.



DR WPI; 2002-221707/28.  
XX P-PSDB; ABB06244.  
PT Reelin protein CR-50 epitope region, useful for diagnosis and treatment  
PT of cerebral disturbance  
XX  
PS Claim 5; Page 11; 16pp; Japanese.  
XX  
CC The present sequence encodes the mouse reelin protein CR-50 epitope  
CC region, which contains the CR-50 antibody recognition site and is free  
CC from F-spondin domains and repetitive sites. Also described are: (1) an  
CC expression vector comprising a polynucleotide encoding a reelin protein  
CC epitope region; (2) host cells with transfected the expression vector;  
CC (3) polypeptides prepared by culture of the host cells; and (4)  
CC polynucleotides comprising the 351 base sequence given in ABL40165 which  
CC encodes the 117 amino acid sequence given in ABB06244; and (5) use of  
CC the polynucleotide for diagnosis and/or treatment of diseases caused by  
CC abnormal positioning of neural cells, and stimulation of association of  
CC reelin protein. The mouse reelin protein CR-50 epitope region has  
CC neuroprotective activity, and can be used in the diagnosis and treatment  
CC of cerebral disturbance due to an abnormal reelin gene and positioning  
CC of neurons.

XX SQ Sequence 351 BP; 86 A; 98 C; 92 G; 75 T; 0 other;  
Query Match 100.0%; Score 351; DB 24; Length 351;  
Best Local Similarity 100.0%; Pred. No. 1.6e-166;  
Matches 351; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GAGCAGTGTGGCACCACATCATGTCGCAATGCTGTACCTTCTGTGAGCGGTACGGCCCT 60  
Db 1 GAGCAGTGTGGCACCACATCATGTCGCAATGCTGTACCTTCTGTGAGCGGTACGGCCCT 60  
Qy 61 CGAGAGTGTGGCACCACATCATGTCGCAATGCTGTACCTTCTGTGAGCGGTACGGCCCT 120  
Db 61 CGAGAGTGTGGCACCACATCATGTCGCAATGCTGTACCTTCTGTGAGCGGTACGGCCCT 120  
Qy 121 GGGTCAGGATCATGTCGATTTAGTTACTTCTGACCCAGCATCTGTCTATACGCCAAG 180  
Db 121 GGGTCAGGATCATGTCGATTTAGTTACTTCTGACCCAGCATCTGTCTATACGCCAAG 180  
Qy 181 AACATACCGCTGATGGATTGAGTTCAGTGGAGAAATTTAGAGCCCTTCCAAATGTGAGCACA 240  
Db 181 AACATACCGCTGATGGATTGAGTTCAGTGGAGAAATTTAGAGCCCTTCCAAATGTGAGCACA 240  
Qy 241 GTCATCCACATCTGTACCTCCCGAGGAGAACCAAGGGGAGAGCGGTTCAGTGG 300  
Db 241 GTCATCCACATCTGTACCTCCCGAGGAGAACCAAGGGGAGAGCGGTTCAGTGG 300  
Qy 301 AACAGGAGACGCTGCGAGTGGGTGAGTGTATGAGGCGCTGCTGGGCCCTG 351  
Db 301 AACAGGAGACGCTGCGAGTGGGTGAGTGTATGAGGCGCTGCTGGGCCCTG 351

RESULT 2  
AAD22754  
ID AAD22754 standard; cDNA; 2745 BP.  
XX  
AC AAD22754;  
XX  
DT 26-FEB-2002 (first entry)  
XX  
DE Mus musculus truncated reelin cDNA.  
XX  
KW Mouse; reelin; F-spondin domain; CR-50 epitope; gene therapy; agyria;  
KW polymicrogyria; ectopic gray matter; ss.  
XX  
OS Mus musculus.  
XX  
FH Key Location/Qualifiers  
FT CDS 283..2052  
FT /\*tag= a  
FT /product= "Mouse truncated reelin protein"

FT sig\_peptide 283..363  
FT /\*tag= b  
FT mat\_peptide 364..2049  
FT /\*tag= c  
FT /product= "Mature truncated reelin protein"  
FT 284..849  
FT /\*tag= d  
FT /note= "Encodes F-spondin domain"  
FT 970..1320  
FT /\*tag= e  
FT /note= "Encodes CR-50 epitope region"  
XX  
XX EP1149844-A2.  
XX 31-OCT-2001.  
XX  
XX 11-APR-2001; 2001BP-0303411.  
XX  
PR 11-APR-2000; 2000JP-0109954.  
XX  
XX (PKE) RIKEN KK.  
XX  
PI Mikoshiba K, Tabata H, Nakajima K;  
XX  
XX WPI; 2002-019320/03.  
XX P-PSDB; AAE13606.  
XX  
PT Novel truncated Reelin protein containing F-spondin domain and CR-50  
PT recognition site of Reelin protein, but not having Reelin repeat site,  
PT useful to treat diseases including agyria due to abnormal neuron  
PT alignment  
XX  
PS Claim 10; Page 20-26; 47pp; English.  
XX  
CC The invention relates to a truncated Reelin protein comprising a  
CC F-spondin domain and a CR-50 recognition site but no reelin protein  
CC repeat site. Reelin is an essential molecule in developing a normal  
CC laminated structure of cerebrum. The truncated reelin protein and its  
CC DNA are useful for treating diseases including agyria, polymicrogyria,  
CC and ectopic gray matter due to abnormal neuronal alignment. Truncated  
CC reelin protein DNA is useful in gene therapy. The present sequence is  
CC a cDNA encoding Mus musculus truncated reelin protein.

SQ Sequence 2745 BP; 661 A; 716 C; 714 G; 654 T; 0 other;  
Query Match 100.0%; Score 351; DB 24; Length 2745;  
Best Local Similarity 100.0%; Pred. No. 1.5e-166;  
Matches 351; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GAGCAGTGTGGCACCACATCATGTCGCAATGCTGTACCTTCTGTGAGCGGTACGGCCCT 60  
Db 970 GAGCAGTGTGGCACCACATCATGTCGCAATGCTGTACCTTCTGTGAGCGGTACGGCCCT 1029  
Qy 61 CGAGAGTGTGGCACCACATCATGTCGCAATGCTGTACCTTCTGTGAGCGGTACGGCCCT 120  
Db 1030 CGAGAGTGTGGCACCACATCATGTCGCAATGCTGTACCTTCTGTGAGCGGTACGGCCCT 1089  
Qy 121 GGGTCAGGATCATGTCGATTTAGTTACTTCTGACCCAGCATCTGTCTATACGCCAAG 180  
Db 1090 GGGTCAGGATCATGTCGATTTAGTTACTTCTGACCCAGCATCTGTCTATACGCCAAG 1149  
Qy 181 AACATACCGCTGATGGATTGAGTTCAGTGGAGAAATTTAGAGCCCTTCCAAATGTGAGCACA 240  
Db 1150 AACATACCGCTGATGGATTGAGTTCAGTGGAGAAATTTAGAGCCCTTCCAAATGTGAGCACA 1209  
Qy 241 GTCATCCACATCTGTACCTCCCGAGGAGAACCAAGGGGAGAGCGGTTCAGTGG 300  
Db 1210 GTCATCCACATCTGTACCTCCCGAGGAGAACCAAGGGGAGAGCGGTTCAGTGG 1369  
Qy 301 AACAGGAGACGCTGCGAGTGGGTGAGTGTATGAGGCGCTGCTGGGCCCTG 351  
Db 1270 AACAGGAGACGCTGCGAGTGGGTGAGTGTATGAGGCGCTGCTGGGCCCTG 1320

US Designate

**THIS PAGE BLANK (USPTO)**



RESULT 3

ABA92603  
ID ABA92603 standard; cDNA; 11673 BP.

XX AC ABA92603;  
XX DT 21-MAR-2002 (first entry)

XX DE Mouse reelin encoding cDNA SEQ ID NO:3.

XX KW Mouse; reelin; low density lipoprotein receptor; LDLR; neuroprotective;  
XX KW extracellular glycoprotein; nontropic; antilipemic; Alzheimer's disease;  
XX KW neurodegenerative disorder; neuronal regeneration; cognitive function;  
XX KW lipid metabolism disease; memory; developmental disorder; gene; ss.

XX OS Mus musculus.

XX FH Key Location/Qualifiers  
XX FT CDS 283..10668  
XX FT /tag= a  
XX FT /product= "mouse reelin"

XX PN US6323177-B1.

XX PD 27-NOV-2001.

XX PF 16-JUN-1999; 99US-0334220.

XX PR 16-JUN-1999; 99US-0334220.

XX PA (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.

XX PI Curran T, D'Arcangelo G;

XX DR WPI; 2002-096596/13.

XX DR P-PSDB; ABB05008.

XX PT Novel composition useful for screening compounds that modulate Reelin  
XX PT binding to low density lipoprotein receptor, comprising an isolated  
XX PT Reelin polypeptide and low density lipoprotein receptor

XX PS Example 1; Column 65-74; 45pp; English.

XX CC The present invention describes a composition (I) comprising an  
XX CC isolated reelin protein (II) bound to an isolated low density lipoprotein  
XX CC receptor (LDLR) (III). (II) is an extracellular glycoprotein of  
XX CC approximately 385 kDa containing a small region of similarity with  
XX CC E-spondin at the N terminus, a stretch of positively charged amino  
XX CC acids at the C terminus, and a series of eight internal repeats of  
XX CC 350-390 amino acids, each repeat containing two related sub-domains  
XX CC that flank a pattern of conserved cysteine residues known as an  
XX CC epidermal growth factor (EGF)-like motif. (I) has neuroprotective,  
XX CC nontropic and antilipemic activities, and can be used as a modulator  
XX CC of reelin-LDLR interaction. (II) is useful in screen for compounds that  
XX CC modulate reelin binding to an LDLR, in an assay system, where the assay  
XX CC system comprises a microplate array and an automated robotic  
XX CC microprocessor controlled system for adding and removing reagents to  
XX CC the microplate array. The compounds identified by the above screening  
XX CC method are useful as therapeutic agents to provide or alleviate a  
XX CC diverse spectrum of diseases including neurodegenerative disorders such  
XX CC as Alzheimer's disease, to facilitate neuronal regeneration after  
XX CC injury, to prevent or alleviate lipid metabolism diseases, to enhance  
XX CC cognitive functions and memory or to ameliorate other developmental  
XX CC disorders. The present sequence encodes mouse (Mus musculus) reelin,  
XX CC which is used in the exemplification of the present invention.

XX SQ Sequence 11673 BP; 2831 A; 2985 C; 2985 G; 2872 T; 0 other;

Query Match 100.0%; Score 351; DB 24; Length 11673;

Best Local Similarity 100.0%; Pred. No. 1.4e-166;

Matches 351; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGCAGTGTGGCACCACCATCATGCAATGCTGCTCACCTTCTGTGAGCGCTAGCGCCT 60  
DB 970 GAGCAGTGTGGCACCACCATCATGCAATGCTGCTCACCTTCTGTGAGCGCTAGCGCCT 1029  
QY 61 CGAGAGCTGACCCACCATGCGCTGAACACAAACAGCATCTGCTCTCCAGTTTTCATT 120  
DB 1030 CGAGAGCTGACCCACCATGCGCTGAACACAAACAGCATCTGCTCTCCAGTTTTCATT 1089  
QY 121 GGTCAGGATCATGTGATTTAGTTTACTCTGACCCAGCATCATCTGTGTATACGCAAG 180  
DB 1090 GGTCAGGATCATGTGATTTAGTTTACTCTGACCCAGCATCATCTGTGTATACGCAAG 1149  
QY 181 AACAAATACCGCTGATGGATTCAGCTGGAGAAAATAGAGCCCTTCCCAATGTGAGCACA 240  
DB 1150 AACAAATACCGCTGATGGATTCAGCTGGAGAAAATAGAGCCCTTCCCAATGTGAGCACA 1209  
QY 241 GTCATCCACATCTGTACCTCCCGAGGAGCAACAGGAGAGCGTGCAAGTCCAGTGG 300  
DB 1210 GTCATCCACATCTGTACCTCCCGAGGAGCAACAGGAGAGCGTGCAAGTCCAGTGG 1269  
QY 301 AACAGGACAGCGCTGCGAGTGGTGAGGTGATGAGCGCTGTGGGCCCTG 351  
DB 1270 AACAGGACAGCGCTGCGAGTGGTGAGGTGATGAGCGCTGTGGGCCCTG 1320

RESULT 4

ABI99284

ID ABI99284 standard; cDNA; 11673 BP.

XX AC ABI99284;

XX DT 07-MAR-2002 (first entry)

XX DE Mouse ischaemic condition related cDNA sequence SEQ ID NO:128.

XX KW Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;

XX KW vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.

XX OS Mus musculus.

XX PN WO2001188188-A2

XX PD 22-NOV-2001.

XX PF 18-MAY-2001; 2001WO-JP04192.

XX PR 18-MAY-2000; 2000JP-0145977.

XX PA (UYN1-) UNIV NIHON SCHOOL JURIDICAL PERSON.

XX PI Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;

XX DR WPI; 2002-034733/04.

XX DR P-PSDB; ABB57065.

XX PT Examining the ischemic condition (e.g. occlusive ischemia) by measuring  
XX PT expression levels of particular genes defined in the specification or  
XX PT by determining the expression profile of a gene group comprising these  
XX PT genes

XX PS Claim 2; Page 362-385; 2690pp; English.

XX CC The present invention describes a method for examining ischaemic  
XX CC conditions, comprising measuring the expression levels of particular  
XX CC genes (I) in a test sample or determining the expression profile of a  
XX CC gene group in the sample comprising genes selected from (I). The method  
XX CC is useful for examining the ischaemic condition (e.g. compressive  
XX CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring  
XX CC expression levels of particular genes (ABI99202 to ABI99912, encoding  
XX CC the protein sequences in ABB57020 to ABB57374) or by determining the  
XX CC expression profile of a gene group comprising these genes. The  
XX CC expression levels or expression profiles produced by these genes are  
XX CC used as an indicator when screening for ischaemic condition-improving

**THIS PAGE BLANK (USPTO)**

progression of liver cancer, hepatocellular carcinoma or metastatic liver tumour in a patient, and differentiating metastatic liver cancer from hepatocellular carcinoma in a patient, involving detecting the level of expression of two or more genes represented in ABN93503-ABN97455 in a tissue sample. The method of the invention has hepatotropic, and cytostatic activity. The method is useful for diagnosing and detecting the progression of liver cancer, hepatocellular carcinoma and metastatic liver carcinoma in a patient. The method is useful for identifying expression profiles which serve as useful diagnostic markers as well as markers that can be used to monitor disease states, disease progression, drug toxicity, drug efficacy and drug metabolism.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [http://www.wipo.int/pub/published\\_pct\\_sequences](http://www.wipo.int/pub/published_pct_sequences).

```

SQ Sequence 11580 BP; 3014 A; 2696 C; 2753 G; 3116 T; 1 other;
      Query Match      7.4%; Score 26; DB 24; Length 11580;
      Best Local Similarity 100.0%; Pred. No. 0.0041;
      Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

DE Human reelin encoding cDNA SEQ ID NO:4.

XX

PN US6323177-B1

27 - NOV - 2001

XX  
PF 16-JUN-1999; 99US-0334220.

XX  
PR 16-JUN-1999: 99US-0334220.

PA (STUD-) ST TURE CHIDREN/C DES HOCENTRY  
XX 16 JUN 1955) 3000 0334220.

PA (SJOUD-) ST JUDE CHILDREN'S  
XX

PI Curran T, D'Arcangelo  
XX

DR WPI; 2002-096596/13.  
DR P-PSDB: ABR05007

F-PDSB; ABB03007.

PT Novel composition useful for screening compounds that modulate Reelin binding to low density lipoprotein receptor, comprising an isolated

Reelin polypeptide and low density lipoprotein receptor

Example 1; Column 75-84; 45pp; English.

XX  
CC The present invention describes a composition (I) comprising an

isolated reelin protein (II) bound to an isolated low density lipoprotein receptor (LDLR) (III). (II) is an extracellular glycoprotein of

approximately 385 kDa containing a small region of similarity with

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 16:41:41 ; Search time: 2213 Seconds  
(without alignments)  
2568.737 Million cell updates/sec

Title: US-09-897-438B-1  
Perfect score: 351  
Sequence: 1 gagcagtggtgaccatcat.....atgaggcctgctggccctg 351

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:\*

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estnu:\*
- 4: em\_estnu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_estro:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_est3:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: gb\_gss:\*
- 18: em\_gss\_hum:\*
- 19: em\_gss\_inv:\*
- 20: em\_gss\_pln:\*
- 21: em\_gss\_vrt:\*
- 22: em\_gss\_fun:\*
- 23: em\_gss\_mam:\*
- 24: em\_gss\_mus:\*
- 25: em\_gss\_other:\*
- 26: em\_gss\_pro:\*
- 27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	ID	Description
C 1	134	38.2	600	12	BG803882
C 2	94	26.8	635	10	BB248113
C 3	26	7.4	315	12	BE938667
C 4	22	6.3	568	12	BF387628
C 5	20	5.7	431	10	BB819780
C 6	19	5.4	404	14	H58973

C 7	19	5.4	416	9	AI400688
C 8	19	5.4	525	10	AW178252
C 9	19	5.4	539	10	AW503344
C 10	19	5.4	741	10	BE288244
C 11	19	5.4	784	13	BM051508
C 12	19	5.4	845	12	BE974073
C 13	19	5.4	899	12	BF680728
C 14	19	5.4	1007	12	BF680728
C 15	19	5.4	1575	13	BM480049
C 16	19	5.4	1575	13	BM480049
C 17	18	5.1	125	17	BH779420
C 18	18	5.1	164	17	BH124575
C 19	18	5.1	187	10	BB118397
C 20	18	5.1	304	10	BB605282
C 21	18	5.1	306	17	AZ765799
C 22	18	5.1	320	12	BF748992
C 23	18	5.1	377	17	AZ816986
C 24	18	5.1	379	9	AA068201
C 25	18	5.1	379	9	AA583727
C 26	18	5.1	384	9	AA794374
C 27	18	5.1	395	17	AZ106948
C 28	18	5.1	399	14	BQ329315
C 29	18	5.1	414	9	AA502373
C 30	18	5.1	449	12	BF706216
C 31	18	5.1	452	17	BH775447
C 32	18	5.1	458	14	BQ335482
C 33	18	5.1	475	12	BF390018
C 34	18	5.1	477	10	AV918544
C 35	18	5.1	480	17	BH782630
C 36	18	5.1	488	17	BH784939
C 37	18	5.1	490	17	BH882572
C 38	18	5.1	493	17	AQ215835
C 39	18	5.1	502	17	BH873530
C 40	18	5.1	512	12	BG737680
C 41	18	5.1	516	10	AV947184
C 42	18	5.1	548	12	BF831988
C 43	18	5.1	574	13	BJ482511
C 44	18	5.1	582	17	BH873271
C 45	18	5.1	627	12	BG446777

## ALIGNMENTS

RESULT 1  
BG803882/c  
LOCUS 0243-03 Mouse El4.5 retina lambda ZAP II Library Mus musculus cDNA, EST 20-DEC-2001  
DEFINITION mRNA sequence.  
ACCESSION BG803882.1 GI:17950794  
VERSION BG803882.1  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
REFERENCE 1 (bases 1 to 600)  
AUTHORS Mu, X., Zhao, S., Pershad, R., Hsieh, T.-F., Scarpa, A., Wang, S.W., White, R.A., Beremand, P.D., Thomas, T.L., Gan, L., and Klein, W.H.  
TITLE Gene expression in the developing mouse retina by EST sequencing and microarray analysis  
JOURNAL Nucleic Acids Res. 29 (24), 4983-4993 (2001)  
MEDLINE 21671825  
COMMENT Contact: Klein WH  
Department of Biochemistry and Molecular Biology  
University of Texas M.D. Anderson Cancer Center  
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA  
Tel: 713 792 3646  
Fax: 713 790 0329.  
Location/Qualifiers  
1..600  
/organism="Mus musculus"  
/db\_xref="taxon:10090"

```

/clone_lib="Mouse E14.5 retina lambda ZAP II Library"
/tissue_type="neural retina"
/dev_stage="embryonic day 14.5 post-fertilization"
/notes="Vector: pAMP10 (Gibco); Cloned unidirectionally.
Primer: Oligo dt. RNA Isolation: cytoplasmic RNA preps
(Mannatis); Cloning Technique: CUA Cloning (CloneAmp,
Life Technologies); Average insert size: 1.8 kb;
Insertion site: TACGTCCACTGAATCTGAGTC-->. Other
information regarding entire library may be found at
http://ppa.swmed.edu/bata/Libraries/microarray_cdna_librar
ies.htm."
BASE COUNT      127 a  167 c  152 g  152 t      2 others
ORIGIN
Query Match      38.2%; Score 134; DB 12; Length 600;
Best Local Similarity 100.0%; Pred. No. 2.5e-62;
Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 218 GAGCCCTTCCAATGTGAGCACAGTCATCCACATCTCTGTACCTCCCGAGGAAGCCAAAG 277
Db 476 GAGCCCTTCCAATGTGAGCACAGTCATCCACATCTCTGTACCTCCCGAGGAAGCCAAAG 417
QY 278 GGGAGAGCGTGCAGTTCACGTGGAACAGACAGCCCTGCCAGTGGGTGAGGTGTATCAGG 337
Db 416 GGGAGAGCGTGCAGTTCACGTGGAACAGACAGCCCTGCCAGTGGGTGAGGTGTATCAGG 357
QY 338 CCTGCTGGGCCCTG 351
Db 356 CCTGCTGGGCCCTG 343

RESULT 2
BB248113
LOCUS
DEFINITION
BB248113 RIKEN full-length enriched, 7 days neonate cerebellum Mus
musculus cDNA clone A730023J04 3', mRNA sequence.
BB248113
ACCESSION
VERSION
KEYWORDS
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 635)
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Kouda
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)
On Jul 6, 2000 this sequence version replaced gi:8940859.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format

```

```

sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
FEATURES
Location/Qualifiers
1..635
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="A730023J04"
/clone_lib="RIKEN full-length enriched, 7 days neonate
cerebellum"
/tissue_type="cerebellum"
/dev_stage="7 days neonate"
/lab_host="DH10B"
/notes="Site_1: SalI; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGGATCCAGAGCTCTTTTCTTTTNN 3'], cDNA was
prepared by using trihalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGAGATTCGAGTTAATTAATTCCTCCCTCCCTCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I."
BASE COUNT      182 a  122 c  118 g  213 t
ORIGIN
Query Match      26.8%; Score 94; DB 10; Length 635;
Best Local Similarity 100.0%; Pred. No. 2.3e-40;
Matches 94; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 29 ATGCTGTACCTTCTGTGAGCGCTACGGCCCTCGAGAGTGCACACATGCTGAACA 88
Db 1 ATGCTGTACCTTCTGTGAGCGCTACGGCCCTCGAGAGTGCACACATGCTGAACA 60
QY 89 CAACAACAGCATCTGCTCCAGTTTTCATTGG 122
Db 61 CAACAACAGCATCTGCTCCAGTTTTCATTGG 94

RESULT 3
BB938667/c
LOCUS
DEFINITION
BB938667 QV0-TN0084-180800-342-a08 TN0084 Homo sapiens cDNA, mRNA sequence.
ACCESSION
VERSION
BB938667.1 GI:10466224
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

REFERENCE  
AUTHORS

1 (bases 1 to 315)  
Dias Neto, E., Garcia Correa, R., Verjovsky-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

## TITLE

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL  
MEDLINE  
COMMENT

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

TELEPHONE

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
{http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-QV0-TN0084-180  
800-342-a086t3-2000-08-18&t4=1}

Seq primer: puc 18 forward

High quality sequence stop: 315.

## FEATURES

source

1. 315  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="TN0084"  
/dev\_stage="Adult"  
/note="Organ: testis\_normal; Vector: puc18; Site\_1: Sma1;  
Site\_2: Sma1; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
84 a 54 c 78 g 89 t

BASE COUNT  
ORIGIN

Query Match 7.4%; Score 26; DB 12; Length 315;  
Best Local Similarity 100.0%; Pred. No. 0.004;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GAGAAATTAGAGCCCTTCCAAATGT 233

|||||

Db 178 GAGAAATTAGAGCCCTTCCAAATGT 153

RESULT 4  
LOCUS

BF387628/c 568 bp mRNA linear EST 27-NOV-2000  
UI-R-CAL-bbs-e-06-0-UI.s1 UI-R-CAL Rattus norvegicus cDNA clone  
UI-R-CAL-bbs-e-06-0-UI 3', mRNA sequence.

## ACCESSION

BF387628

VERSION

BF387628.1

KEYWORDS

EST.

SOURCE

ORGANISM

Norway rat.

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 568)

REFERENCE

AUTHORS

Bonaldi, M.F., Lennon, G., and Soares, M.B.

TITLE

Normalization and subtraction: Two approaches to facilitate gene

discovery

JOURNAL

Genome Res. 6 (9), 791-806 (1996)

MEDLINE

9704447

COMMENT

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250  
Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized pons library cDNA library preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)  
Seq primer: M13 Forward  
POLYA=Yes.

FEATURES  
source

Location/Qualifiers

1. 568

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db\_xref="taxon:10116"

/clone="UI-R-CAL-bbs-e-06-0-UI"

/clone\_lib="UI-R-CAL"

/lab\_host="DHI0B (Life Technologies)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-CAL library is a subtracted library derived from the following tissues: thalamus, cerebellum, hypothalamus, medulla, pons, midbrain, cerebral cortex, corpus striatum, testis, and hippocampus. For a detailed description of the library from which this clone was derived, please visit our web site at [ratest.eng.uiowa.edu](http://ratest.eng.uiowa.edu). The subtraction has been previously described in (Bonaldi, Lennon and Soares, Genome Research 6:791-806, 1996)

TAG\_LTB=UI-R-CAL

TAG\_TISSUE=pons

TAG\_SEQ=AGCAGC

BASE COUNT 116 a 156 c 131 g 165 t

## ORIGIN

Query Match 6.3%; Score 22; DB 12; Length 568;

Best Local Similarity 100.0%; Pred. No. 0.75;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 289 CAGTTCAGTGGAAACAGGACA 310

|||||

Db 550 CAGTTCAGTGGAAACAGGACA 529

## RESULT 5

LOCUS

BB819780

DEFINITION

BB819780 431 bp mRNA linear EST 19-NOV-2001  
BB819780 RIKEN full-length enriched, mammary gland RCB-0526  
Jyg-MC(A) cDNA Mus musculus cDNA clone G830002B05 3', mRNA  
sequence.

ACCESSION

BB819780

VERSION

BB819780.1

KEYWORDS

house mouse.

SOURCE

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 431)

REFERENCE

AUTHORS

Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T.,

Hayashizaki, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T.,

Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T.,

Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sakaki, D., Sato, K.,

Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa,

A., Takahashi, F., Takaku-Akashira, S., Tanaka, T., Tomaru, A., Toya, T.,

Watahiki, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al.

2001)

TITLE

JOURNAL

Unpublished (2001)

COMMENT

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsr.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh  
M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
Wataniki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura  
S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and  
Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara  
Y. and Hayashizaki, Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (http://genome.gsc.riken.go.jp) for  
further details.  
e mouse tissues.

#### FEATURES source

1. .431  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="G830002B05"  
/clone\_lib="RIKEN full-length enriched, mammary gland  
RCB-0526 Jyg-MC(A) cDNA"  
/tissue\_type="mammary gland"  
/cell\_line="RCB-0526 Jyg-MC(A)"  
/note="pooled cell lines; (cell\_line=CRL-1751 WEHI 164),  
(cell\_line=CRL-2116 JC), (cell\_line=RCB-0035 WEHI-3),  
(cell\_line=RCB-0464 Meth-A), (cell\_line=RCB-0545 OHTA),  
(cell\_line=RCB-0559 K-1 F1), (cell\_line=RCB-1283 B16  
melanoma), (cell\_type=B cells, cell\_line=CRL-1702 WEHI 231  
), (cell\_type=Teydig cells, cell\_line=CRL-2065 MLTC-1),  
(cell\_type=Nullipotent stem cell, cell\_line=CRL-2070 NE),  
(tissue\_type=bladder, cell\_line=RCB-0544 MBT-2),  
(tissue\_type=bone marrow, cell\_type=stroma cell,  
cell\_line=CRL-2028 SR-4987), (tissue\_type=colon,  
cell\_line=RCB-0549 Cle-H3), (tissue\_type=kidney,  
cell\_line=CCL-142 RAG), (tissue\_type=submandibular gland,  
cell\_line=CRL-1734 SCA-9 clone 15), (strain=BALB/C,  
cell\_type=B cells, cell\_line=CRL-1669 BCL1 Clone 13.20-3B3  
) , (strain=C3H, tissue\_type=brain, cell\_line=CRL-1443  
BC3H1)"

BASE COUNT 113 a 89 c 94 g 135 t  
ORIGIN

Query Match 5.7%; Score 20; DB 10; Length 431;  
Best Local Similarity 100.0%; Pred. No. 8.7;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 205 CTGGAGAAATTAGAGCCCC 224  
|||||  
Db 86 CTGGAGAAATTAGAGCCCC 105

#### RESULT 6

H58973/c  
LOCUS H58973 404 bp mRNA linear EST 06-OCT-1995  
DEFINITION Yr40508.r1 Soares fetal liver spleen INFUS Homo sapiens cDNA clone  
IMAGE:207735 5', mRNA sequence.  
ACCESSION H58973  
VERSION H58973.1 GI:1011805  
KEYWORDS EST.  
SOURCE human.

#### ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 404)  
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman  
M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,  
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisan, E., Waterston  
R., Williamson, A., Wohlmann, P. and Wilson, R.  
The WashU-Merck EST Project  
Unpublished (1995)  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

#### TITLE

High quality sequence stops: 294  
Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 864 Std Error: 0.00  
Seq primer: M13RP1  
High quality sequence stop: 294.

#### JOURNAL

Location/Qualifiers  
1. .404  
/organism="Homo sapiens"  
/db\_xref="GDB:3776866"  
/db\_xref="taxon:9606"  
/clone="IMAGE:207735"  
/clone\_lib="Soares fetal liver spleen INFUS"  
/sex="male"  
/dev\_stage="20 week-post conception fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)  
with a modified polylinker; Site: 1: Pac I; Site 2: Eco RI;  
1st strand cDNA was primed with a Pac I - oligo(dT) primer  
[5' AACTGGAAGATTAATAAGATCTTTTCTTTTCTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Pac I and cloned into the Pac I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M.Fatima Bonaldo."

#### FEATURES

BASE COUNT 82 a 128 c 88 g 103 t  
ORIGIN  
Query Match 5.4%; Score 19; DB 14; Length 404;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 333 TGAGGCGCTGCTGGCGCTG 351  
|||||  
Db 247 TGAGGCGCTGCTGGCGCTG 229

#### RESULT 7

AI400688/c  
LOCUS AI400688 416 bp mRNA linear EST 08-FEB-1999  
DEFINITION t993e01.x1 NCI\_CGAP\_CLL1 Homo sapiens cDNA clone IMAGE:2116344 3',  
mRNA sequence.  
ACCESSION AI400688  
VERSION AI400688.1 GI:4243775  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 416)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
CONTACT: Robert Strausberg, Ph.D.



Email: cgabbs-r@mail.nih.gov  
 Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever, M.D., Louis M. Staudt, M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.llnl.gov/bbrp/image/image.html  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 342.

## FEATURES

Location/Qualifiers

Source 1..416  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2116344"  
 /clone\_lib="NCI\_CGAP\_CLL1"  
 /tissue\_type="B-cell, chronic lymphocytic leukemia"  
 /lab\_host="DH10B"  
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTCACATCTGAAGTGGGAGCGCGCATTCCTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized, and was constructed by Bento Soares and M. Fatima Bonaldo."  
 BASE COUNT 93 a 92 c 84 g 147 t  
 ORIGIN

Query Match 5.4%; Score 19; DB 9; Length 416;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 286 GTGCAGTTCAGTGGAAC 304

|||||

Db 298 GTGCAGTTCAGTGGAAC 280

## RESULT 8

AW178252/c

LOCUS AW178252 525 bp mRNA linear EST 16-NOV-1999  
 DEFINITION MR0-HT0066-010999-001-e05 HT0066 Homo sapiens cDNA, mRNA sequence.

ACCESSION AW178252

VERSION AW178252.1 GI:6444289

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 525)

HCGP http://www.ludwig.org.br/ORESTES.

The FAPESP/LICR Human Cancer Genome Project

Unpublished (1999)

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR0&t2=MR0-HT0066-010999-001-e05&t3=1999-09-01&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 40

High quality sequence stop: 99.

## FEATURES

Location/Qualifiers

1..525

/organism="Homo sapiens"

## SOURCE

/db\_xref="taxon:9606"  
 /clone\_lib="HT0066"  
 /dev\_stage="Adult"  
 /note="Organ: head\_neck; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."  
 BASE COUNT 247 a 57 c 119 g 102 t  
 ORIGIN

Query Match 5.4%; Score 19; DB 10; Length 525;  
 Best Local Similarity 100.0%; Pred. No. 33;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 TCCTGCTCCAGTTTCCA 118

|||||

Db 126 TCCTGCTCCAGTTTCCA 108

## RESULT 9

AW503344

LOCUS

DEFINITION AW503344 539 bp mRNA linear EST 01-MAR-2000  
 UI-RF-BN0-akx-f-10-0-UI-r1 NIH\_MGC\_50 Homo sapiens cDNA clone

ACCESSION AW503344

VERSION AW503344.1 GI:7118645

KEYWORDS EST.

SOURCE human.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 539)

NIH-MGC http://mgc.nci.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs-r@mail.nih.gov

Eco RI site shown at the beginning of the sequence.

Tissue Procurement: Louis M. Staudt, M.D., Ph.D.

cDNA Library Preparation: M.B. Soares Lab

cDNA Library Arrayed by: M.B. Soares Lab

DNA Sequencing by: M.B. Soares Lab

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Seq primer: M13 Forward

Location/Qualifiers

1..539

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:3078690"

/clone\_lib="NIH\_MGC\_50"

/tissue\_type="lymph"

/cell\_type="germinal center B cells"

/cell\_line="MGC85"

/lab\_host="DH10B (LTI)"

/note="Vector: pT7T3-Pac; Site\_1: NotI; Site\_2: Eco RI;

Constructed from size fractionated cytoplasmic mRNA

(3.5-4.4Kb). Directionally cloned. Cells provided by

Louis M. Staudt, Ph.D. Library preparation by Maria de

Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D."

BASE COUNT 158 a 126 c 108 g 147 t

## ORIGIN

Query Match 5.4%; Score 19; DB 10; Length 539;

Best Local Similarity 100.0%; Pred. No. 33;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 286 GTGCAGTTCAGTGGAAC 304

|||||

```

Db 412 GTGAGTTCACGTGGAAAC 430

RESULT 10
BE288244/c
LOCUS 601095415f1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3489986 5',
DEFINITION mRNA sequence.
ACCESSION BE288244
VERSION BE288244.1 GI:9167806
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 741)
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="NCI_CGAP_Mam5"
/tissue_type="tumor, gross tissue"
/lab_host="DH10B"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,
NIH"
BASE COUNT 142 a 214 c 189 g 193 t 3 others
ORIGIN

Query Match 5.4%; Score 19; DB 10; Length 741;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 286 GTGAGTTCACGTGGAAAC 304
|||||
Db 265 GTGAGTTCACGTGGAAAC 247

RESULT 11
BM051508
LOCUS 603638192f1 NTH_MGC_8 Homo sapiens cDNA clone IMAGE:5419438 5',
DEFINITION mRNA sequence.
ACCESSION BM051508
VERSION BM051508.1 GI:16780775
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 784)
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_8"
* TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

Db 412 GTGAGTTCACGTGGAAAC 430

RESULT 10
BE288244/c
LOCUS 601095415f1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3489986 5',
DEFINITION mRNA sequence.
ACCESSION BE288244
VERSION BE288244.1 GI:9167806
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 741)
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="NCI_CGAP_Mam5"
/tissue_type="tumor, gross tissue"
/lab_host="DH10B"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Robin Humphreys,
NIH"
BASE COUNT 142 a 214 c 189 g 193 t 3 others
ORIGIN

Query Match 5.4%; Score 19; DB 10; Length 741;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 286 GTGAGTTCACGTGGAAAC 304
|||||
Db 265 GTGAGTTCACGTGGAAAC 247

RESULT 11
BM051508
LOCUS 603638192f1 NTH_MGC_8 Homo sapiens cDNA clone IMAGE:5419438 5',
DEFINITION mRNA sequence.
ACCESSION BM051508
VERSION BM051508.1 GI:16780775
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 784)
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_8"
* TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1872 row: g column: 23
High quality sequence stop: 748.
Location/Qualifiers
1. .784
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5419438"
/clone_lib="NIH_MGC_8"
/tissue_type="Burkitt lymphoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 256 a 170 c 168 g 190 t
ORIGIN

Query Match 5.4%; Score 19; DB 13; Length 784;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 286 GTGAGTTCACGTGGAAAC 304
|||||
Db 220 GTGAGTTCACGTGGAAAC 238

RESULT 12
BE974073/c
LOCUS 601680409f1 NTH_MGC_83 Homo sapiens cDNA clone IMAGE:3950499 5',
DEFINITION mRNA sequence.
ACCESSION BE974073
VERSION BE974073.1 GI:10587409
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 845)
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3950499"
/clone_lib="NIH_MGC_83"
* TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM817 row: b column: 04
High quality sequence stop: 567.
Location/Qualifiers
1. .845
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3950499"
/clone_lib="NIH_MGC_83"

```

/lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: prostate; Vector: pDNR-LIB (Clontech);  
 Site\_1: SfII (ggcgcctcgcc); Site\_2: SfII (ggccattatggcc  
 ); 5' and 3' adaptors were used in cloning as follows: 5'  
 adaptor sequence: 5'-CAGGCCATTATGGCC-3' and 3' adaptor  
 sequence: 5'-ATTCTAGAGCGCGGCGGCACATG-dt(30)BN-3'  
 (where B = A, C, G, or T). Average  
 insert size 1.4 kb (range 0.5-4.0 kb). 14/15 colonies  
 contained inserts by PCR. This library was enriched for  
 full-length clones and was constructed by Clontech  
 Laboratories (Palo Alto, CA)."

BASE COUNT 334 a 89 c 224 g 198 t

ORIGIN

Query Match 5.4%; Score 19; DB 12; Length 845;  
 Best Local Similarity 100.0%; Pred. No. 37;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 TCTGTCTCTCCAGTTTCCA 118  
 |||||

Db 613 TCTGTCTCTCCAGTTTCCA 595

RESULT 13  
 BF680728/c  
 LOCUS  
 DEFINITION 899 bp mRNA linear EST 21-DEC-2000  
 602155347F1 NIH\_MGC\_83 Homo sapiens cDNA clone IMAGE:4296395 5',  
 mRNA sequence.

ACCESSION  
 VERSION BF680728  
 KEYWORDS  
 SOURCE EST.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS 1 (bases 1 to 899)  
 TITLE NIH-MGC http://mgc.nci.nih.gov/  
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
 COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.  
 Email: cyapbs-r@mail.nih.gov  
 Tissue Procurement: CLONTECH Laboratories, Inc.  
 CDNA Library Preparation: CLONTECH Laboratories, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov

Plate: LLCM148 row: n column: 12  
 High quality sequence stop: 615.

Location/Qualifiers

## FEATURES

source

1..899  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4296395"  
 /clone\_lib="NIH\_MGC\_83"  
 /lab\_host="DH10B (T1 phage-resistant)"

/note="Organ: prostate; Vector: pDNR-LIB (Clontech);  
 Site\_1: SfII (ggcgcctcgcc); Site\_2: SfII (ggccattatggcc  
 ); 5' and 3' adaptors were used in cloning as follows: 5'  
 adaptor sequence: 5'-CAGGCCATTATGGCC-3' and 3' adaptor  
 sequence: 5'-ATTCTAGAGCGCGGCGGCACATG-dt(30)BN-3'  
 (where B = A, C, G, or T). Average  
 insert size 1.4 kb (range 0.5-4.0 kb). 14/15 colonies  
 contained inserts by PCR. This library was enriched for  
 full-length clones and was constructed by Clontech  
 Laboratories (Palo Alto, CA)."

BASE COUNT 327 a 139 c 236 g 197 t

ORIGIN

Query Match 5.4%; Score 19; DB 12; Length 899;  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 TCTGTCTCTCCAGTTTCCA 118  
 |||||

Db 633 TCTGTCTCTCCAGTTTCCA 615

RESULT 14  
 BG484905/c

LOCUS  
 DEFINITION 1007 bp mRNA linear EST 21-MAR-2001  
 602503910F1 NIH\_MGC\_77 Homo sapiens cDNA clone IMAGE:4617170 5',  
 mRNA sequence.

ACCESSION  
 VERSION BG484905  
 KEYWORDS  
 SOURCE EST.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS 1 (bases 1 to 1007)  
 TITLE NIH-MGC http://mgc.nci.nih.gov/  
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
 COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.  
 Email: cyapbs-r@mail.nih.gov

Tissue Procurement: CLONTECH Laboratories, Inc.

CDNA Library Preparation: CLONTECH Laboratories, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCM1373 row: d column: 03

High quality sequence stop: 650.

Location/Qualifiers

## FEATURES

source

1..1007  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4617170"  
 /clone\_lib="NIH\_MGC\_77"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site\_1:  
 SfII (ggcgcctcgcc); Site\_2: SfII (ggccattatggcc); 5' and  
 3' adaptors were used in cloning as follows: 5' adaptor  
 sequence: 5'-CAGGCCATTATGGCC-3' and 3' adaptor sequence:  
 5'-ATTCTAGAGCGCGGCGGCACATG-dt(30)BN-3' (where B = A,  
 C, G, or T). Average insert size 1.9  
 kb (range 0.5-4.0 kb). 12/15 colonies contained inserts  
 by PCR. This library was enriched for full-length clones  
 and was constructed by Clontech Laboratories (Palo Alto,  
 CA). Note: this is a NIH\_MGC Library."

BASE COUNT 363 a 182 c 270 g 192 t

ORIGIN

Query Match 5.4%; Score 19; DB 12; Length 1007;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 TCTGTCTCTCCAGTTTCCA 118  
 |||||

Db 640 TCTGTCTCTCCAGTTTCCA 622

RESULT 15  
 BM480049

LOCUS  
 DEFINITION 1575 bp mRNA linear EST 05-FEB-2002  
 AGENCOURT\_6424083 NIH\_MGC\_67 Homo sapiens cDNA clone IMAGE:5491592  
 5', mRNA sequence.

ACCESSION  
 VERSION BM480049  
 KEYWORDS  
 SOURCE EST.

## ORGANISM

human.

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE  
1 (bases 1 to 1575)  
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM12112 row: f column: 09  
High quality sequence stop: 372.

FEATURES  
source  
Location/Qualifiers  
1..1575  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5491592"  
/clone\_lib="NIH\_MGC\_67"  
/tissue\_type="retinoblastoma"  
/lab\_host="DH10B (phage-resistant)"  
/notes="Organ: eye; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.75 kb. Library constructed by Life  
Technologies."  
BASE COUNT 434 a 533 c 270 g 333 t 5 others  
ORIGIN

Query Match 5.4%; Score 19; DB 13; Length 1575;  
Best Local Similarity 100.0%; Pred. No. 45;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 12 CACCATCATGCATGGCAAT 30  
|||||  
Db 734 CACCATCATGCATGGCAAT 752

Search completed: November 6, 2002, 18:11:53  
Job time : 2220 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 17:40:37 ; Search time 89 Seconds  
(without alignments)  
1313.903 Million cell updates/sec

Title: US-09-897-438B-1

Perfect score: 351

Sequence: 1 gagcagctggcaccatcat.....atgagcctgctgggcccctg 351

Scoring table:

OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 310279 seqs, 166577418 residues

Word size : 0

Total number of hits satisfying chosen parameters: 620558

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications, NA.\*

- 1: /cgn2.6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2.6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2.6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2.6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2.6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2.6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2.6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2.6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2.6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 10: /cgn2.6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*
- 11: /cgn2.6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 12: /cgn2.6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*
- 13: /cgn2.6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 14: /cgn2.6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	351	100.0	351	10	US-09-897-438B-1
2	26	7.4	11580	10	US-09-880-107-3436
3	22	6.3	22	10	US-09-897-438B-6
4	21	6.0	21	10	US-09-897-438B-10
5	18	5.1	18	10	US-09-897-438B-11
6	18	5.1	18	10	US-09-880-107-3814
7	17	4.8	325	10	US-09-764-877-2810
8	17	4.8	429	10	US-09-880-107-2470
9	17	4.8	499	10	US-09-864-761-14497
10	17	4.8	559	10	US-09-864-761-9458
11	17	4.8	675	10	US-09-764-864-678
12	17	4.8	1488	10	US-09-758-652-11
13	17	4.8	1732	10	US-09-923-556-1
14	17	4.8	1732	10	US-09-923-556-5
15	17	4.8	2131	10	US-09-764-864-250
16	17	4.8	3400	10	US-09-987-025-1
17	17	4.8	11881	10	US-09-764-869-2124
18	17	4.8	16552	10	US-09-764-855-321
19	17	4.8	16552	10	US-09-764-855-322

20	16	4.6	271	10	US-09-923-876-1444	Sequence 1444, Ap
21	16	4.6	347	10	US-09-833-381-1818	Sequence 1818, Ap
22	16	4.6	347	10	US-09-833-381-1819	Sequence 1819, Ap
23	16	4.6	420	10	US-09-864-761-15602	Sequence 15602, A
24	16	4.6	430	10	US-09-815-343-782	Sequence 782, App
25	16	4.6	445	10	US-09-867-701-10721	Sequence 10721, A
26	16	4.6	465	10	US-09-864-761-29	Sequence 29, Appl
27	16	4.6	494	10	US-09-867-701-1885	Sequence 1885, Ap
28	16	4.6	576	10	US-09-815-343-1004	Sequence 1004, Ap
29	16	4.6	977	10	US-09-821-725-1	Sequence 1, Appl
30	16	4.6	1633	10	US-09-813-358-20	Sequence 20, Appl
31	16	4.6	1690	10	US-09-828-303-2	Sequence 2, Appl
32	16	4.6	1839	10	US-09-828-303-10	Sequence 10, Appl
33	16	4.6	2692	10	US-09-908-322-11	Sequence 11, Appl
34	16	4.6	2857	10	US-09-908-322-4	Sequence 4, Appl
35	16	4.6	3435	10	US-09-803-670-1	Sequence 1, Appl
36	16	4.6	4181	10	US-09-871-916-1	Sequence 1, Appl
37	16	4.6	5588	10	US-09-917-800A-1565	Sequence 1565, Ap
38	16	4.6	8489	10	US-09-821-725-3	Sequence 3, Appl
39	16	4.6	8868	10	US-09-803-670-3	Sequence 3, Appl
40	16	4.6	8948	10	US-09-735-705-119	Sequence 119, App
41	16	4.6	8948	10	US-09-850-716A-119	Sequence 119, App
42	16	4.6	8948	10	US-09-897-778-119	Sequence 119, App
43	16	4.6	9588	10	US-09-954-456-1848	Sequence 1848, Ap
44	16	4.6	10256	10	US-09-764-847-1182	Sequence 1182, Ap
45	16	4.6	13574	10	US-09-070-927A-114	Sequence 114, App

#### ALIGNMENTS

RESULT 1  
US-09-897-438B-1  
; Sequence 1, Application US/09897438B  
; Patent No. US20020137095A1  
; GENERAL INFORMATION:  
; APPLICANT: Mikoshiba, Katsuhiko  
; APPLICANT: Tate, Naoko  
; TITLE OF INVENTION: REELIN PROTEIN CR-50 EPTOPE REGION  
; FILE REFERENCE: 04853-0076-00000  
; CURRENT APPLICATION NUMBER: US/09/897,438B  
; CURRENT FILING DATE: 2001-07-03  
; PRIOR APPLICATION NUMBER: JP 2000-202801  
; PRIOR FILING DATE: 2000-07-04  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 351  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-09-897-438B-1

Query Match	100.0%	Score 351;	DB 10;	Length 351;
Best Local Similarity	100.0%	Pred No. 2.4e-181;		
Matches 351;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	GAGCAGTGTGGCACCACCATCATGCATGCAATGCTGTCACTTCTGTGAGCGGTACGCCCT	60	
Db	1	GAGCAGTGTGGCACCACCATCATGCATGCAATGCTGTCACTTCTGTGAGCGGTACGCCCT	60	
QY	61	CGAGAGCTGACCAACCATCATGCCTGGAACACAAACAGACATCTCTCTCCAGTTTCCATT	120	
Db	61	CGAGAGCTGACCAACCATCATGCCTGGAACACAAACAGACATCTCTCTCCAGTTTCCATT	120	
QY	121	GGTCCAGGATCATGTCGATTAGTTACTCTGACCCAGCATCACTGTGTCTATACGCCAAG	180	
Db	121	GGTCCAGGATCATGTCGATTAGTTACTCTGACCCAGCATCACTGTGTCTATACGCCAAG	180	
QY	181	AACAATACCCCTGATTGGATTTCAGCTGGAGAAAATTAGAGCCCTTCCCAATGTGAGCACA	240	
Db	181	AACAATACCCCTGATTGGATTTCAGCTGGAGAAAATTAGAGCCCTTCCCAATGTGAGCACA	240	
QY	241	GTATCCACATCTGTACCTCCCGGAGAAAGCCAAAGGGAGAGCGTGCAGTTCCAGTGG	300	

Db 241 GTATCACATCCCTGTACCTCCCGAGGAAAGGAGGAGCGTGCAGTTCCAGTGG 300  
QY 301 AAACAGACAGCCTGCGAGTGGGTGAGGTGTATGAGGCGCTGCTGGGCCCTG 351  
Db 301 AAACAGACAGCCTGCGAGTGGGTGAGGTGTATGAGGCGCTGCTGGGCCCTG 351

RESULT 2  
US-09-880-107-3436  
; Sequence 3436, Application US/09880107  
; Patent No. US20020142981A1  
; GENERAL INFORMATION:  
; APPLICANT: Horne, Darci T.  
; APPLICANT: Vockley, Joseph G.  
; APPLICANT: Scherf, Uwe  
; APPLICANT: Gene Logic, Inc.  
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer  
; FILE REFERENCE: 04853-0076-00000  
; CURRENT APPLICATION NUMBER: US/09/880,107  
; CURRENT FILING DATE: 2001-06-14  
; PRIOR APPLICATION NUMBER: US 60/211,379  
; PRIOR FILING DATE: 2000-06-14  
; PRIOR APPLICATION NUMBER: US 60/237,054  
; PRIOR FILING DATE: 2000-10-02  
; NUMBER OF SEQ ID NOS: 3950  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3436  
; LENGTH: 11580  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 U79716  
US-09-880-107-3436

Query Match 7.4%; Score 26; DB 10; Length 11580;  
Best Local Similarity 100.0%; Pred. No. 0.00014;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 208 GAGAAATTAGAGCCCTTCCCAATGT 233  
Db 1067 GAGAAATTAGAGCCCTTCCCAATGT 1092

RESULT 3  
US-09-897-438B-6/c  
; Sequence 6, Application US/09897438B  
; Patent No. US20020137095A1  
; GENERAL INFORMATION:  
; APPLICANT: Mikoshiba, Katsuhiko  
; APPLICANT: Tate, Naoko  
; TITLE OF INVENTION: REELIN PROTEIN CR-50 EPTOPE REGION  
; FILE REFERENCE: 04853-0076-00000  
; CURRENT APPLICATION NUMBER: US/09/897,438B  
; CURRENT FILING DATE: 2001-07-03  
; PRIOR APPLICATION NUMBER: JP 2000-202801  
; PRIOR FILING DATE: 2000-07-04  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 22  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
US-09-897-438B-6

Query Match 6.3%; Score 22; DB 10; Length 22;  
Best Local Similarity 100.0%; Pred. No. 0.016;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 330 GTATGAGCCTGTGGGCCCTG 351

Db 22 GTATGAGCCTGTGGGCCCTG 1  
RESULT 4  
US-09-897-438B-10  
; Sequence 10, Application US/09897438B  
; Patent No. US20020137095A1  
; GENERAL INFORMATION:  
; APPLICANT: Mikoshiba, Katsuhiko  
; APPLICANT: Tate, Naoko  
; TITLE OF INVENTION: REELIN PROTEIN CR-50 EPTOPE REGION  
; FILE REFERENCE: 04853-0076-00000  
; CURRENT APPLICATION NUMBER: US/09/897,438B  
; CURRENT FILING DATE: 2001-07-03  
; PRIOR APPLICATION NUMBER: JP 2000-202801  
; PRIOR FILING DATE: 2000-07-04  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 10  
; LENGTH: 21  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
US-09-897-438B-10

Query Match 6.0%; Score 21; DB 10; Length 21;  
Best Local Similarity 100.0%; Pred. No. 0.055;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GAGCAGTGTGGCACCACCATG 21  
Db 1 GAGCAGTGTGGCACCACCATG 21

RESULT 5  
US-09-897-438B-11  
; Sequence 11, Application US/09897438B  
; Patent No. US20020137095A1  
; GENERAL INFORMATION:  
; APPLICANT: Mikoshiba, Katsuhiko  
; APPLICANT: Tate, Naoko  
; TITLE OF INVENTION: REELIN PROTEIN CR-50 EPTOPE REGION  
; FILE REFERENCE: 04853-0076-00000  
; CURRENT APPLICATION NUMBER: US/09/897,438B  
; CURRENT FILING DATE: 2001-07-03  
; PRIOR APPLICATION NUMBER: JP 2000-202801  
; PRIOR FILING DATE: 2000-07-04  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 11  
; LENGTH: 18  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
US-09-897-438B-11

Query Match 5.1%; Score 18; DB 10; Length 18;  
Best Local Similarity 100.0%; Pred. No. 2.3;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GAGCAGTGTGGCACCACATC 18  
Db 1 GAGCAGTGTGGCACCACATC 18

RESULT 6  
US-09-880-107-3814/c  
; Sequence 3814, Application US/09880107

; Patent No. US20020142981A1  
; GENERAL INFORMATION:  
; APPLICANT: Horne, Darci T.  
; APPLICANT: Vockley, Joseph G.  
; APPLICANT: Scherf, Uwe  
; APPLICANT: Gene Logic, Inc.  
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer  
; FILE REFERENCE: 44921-5028-WO  
; CURRENT APPLICATION NUMBER: US/09/880,107  
; CURRENT FILING DATE: 2001-06-14  
; PRIOR APPLICATION NUMBER: US-60/211,379  
; PRIOR FILING DATE: 2000-06-14  
; (PRIOR APPLICATION NUMBER: US 60/237,054)  
; PRIOR FILING DATE: 2000-10-02  
; NUMBER OF SEQ ID NOS: 3950  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3814  
; LENGTH: 198285  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 X87344  
US-09-880-107-3814

Query Match 5.1%; Score 18; DB 10; Length 198285;  
Best Local Similarity 100.0%; Pred. No. 3.7;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 CTGACCACCATGCGCTG 84  
|||||  
Db 9594 CTGACCACCATGCGCTG 9577

RESULT 7  
US-09-764-877-2810/c  
; Sequence 2810, Application US/09764877  
; Patent No. US20020147140A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC005  
; CURRENT APPLICATION NUMBER: US/09/764,877  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 4031  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2810  
; LENGTH: 325  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-877-2810

Query Match 4.8%; Score 17; DB 10; Length 325;  
Best Local Similarity 100.0%; Pred. No. 9.4;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 GCACCATCATGCGGC 27  
|||||  
Db 165 GCACCATCATGCGGC 149

RESULT 8  
US-09-880-107-2470/c  
; Sequence 2470, Application US/09880107  
; Patent No. US20020142981A1  
; GENERAL INFORMATION:  
; APPLICANT: Horne, Darci T.  
; APPLICANT: Vockley, Joseph G.  
; APPLICANT: Scherf, Uwe  
; APPLICANT: Gene Logic, Inc.  
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer  
; FILE REFERENCE: 44921-5028-WO  
; CURRENT APPLICATION NUMBER: US/09/880,107

; CURRENT FILING DATE: 2001-06-14  
; PRIOR APPLICATION NUMBER: US 60/211,379  
; PRIOR FILING DATE: 2000-06-14  
; PRIOR APPLICATION NUMBER: US 60/237,054  
; PRIOR FILING DATE: 2000-10-02  
; NUMBER OF SEQ ID NOS: 3950  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2470  
; LENGTH: 429  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 N27334  
; NAME/KEY: unsure  
; LOCATION: (1)..(429)  
; OTHER INFORMATION: n = a or c or g or t  
US-09-880-107-2470

Query Match 4.8%; Score 17; DB 10; Length 429;  
Best Local Similarity 100.0%; Pred. No. 9.6;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 335 AGGCCTGCTGGCCCTG 351  
|||||  
Db 276 AGGCCTGCTGGCCCTG 260

RESULT 9  
US-09-864-761-14497/c  
; Sequence 14497, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aeomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21

; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 14497  
; LENGTH: 499  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC000121.1  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.62  
US-09-864-761-14497

Query Match 4.8%; Score 17; DB 10; Length 499;  
Best Local Similarity 100.0%; Pred. No. 9.7; Indels 0; Gaps 0;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 217 AGAGCCCTTCCAATGT 233  
|||||  
Db 487 AGAGCCCTTCCAATGT 471

## RESULT 10

US-09-864-761-9458  
; Sequence 9458, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aeomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; PRIOR FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203

; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 9458  
; LENGTH: 559  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AL022318.2  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.9  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.6  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.2  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.4  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.9  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.7  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.9  
US-09-864-761-9458

Query Match 4.8%; Score 17; DB 10; Length 559;  
Best Local Similarity 100.0%; Pred. No. 9.7; Indels 0; Gaps 0;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 GAGGCTGTGGGCGCT 350  
|||||  
Db 208 GAGGCTGTGGGCGCT 224

## RESULT 11

US-09-764-864-678/c  
; Sequence 678, Application US/09764864  
; Patent No. US20020132753A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT223  
; CURRENT APPLICATION NUMBER: US/09/764,864  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION DATA REMOVED - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1792  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 678  
; LENGTH: 675  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-864-678

Query Match 4.8%; Score 17; DB 10; Length 675;  
Best Local Similarity 100.0%; Pred. No. 9.8; Indels 0; Gaps 0;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 TCTGTCTCCAGTTTC 116  
|||||  
Db 107 TCTGTCTCCAGTTTC 91

## RESULT 12

US-09-758-652-11  
; Sequence 11, Application US/09758652  
; Patent No. US20010011377A1  
; GENERAL INFORMATION:  
; APPLICANT: ANTHONY JOHN KINNEY  
; GARY MICHAEL FADER  
; TITLE OF INVENTION: SUPPRESSION OF SPECIFIC  
; CLASSES OF SOYBEAN SEED  
; PROTEIN GENES  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY  
; STREET: 1007 MARKET STREET  
; CITY: WILMINGTON  
; STATE: DELAWARE  
; COUNTRY: UNITED STATES OF AMERICA



ZIP: 19898  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.50 INCH  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: MICROSOFT WINDOWS 95  
SOFTWARE: MICROSOFT WORD 7.0A  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/758,652  
FILING DATE: 11-Jan-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/019,940  
FILING DATE: JUNE 14, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: LYNNE M. CHRISTENBURY  
REGISTRATION NUMBER: 30,971  
REFERENCE/DOCKET NUMBER: BB-1071-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 302-992-5481  
TELEFAX: 302-773-0164  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1488 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
SEQUENCE DESCRIPTION: SEQ ID NO: 11:  
US-09-758-652-11

Query Match 4.8%; Score 17; DB 10; Length 1488;

Best Local Similarity 100.0%; Pred. No. 10;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 261 CCCCGAGGAGCCCAAG 277  
|||||

Db 898 CCCCGAGGAGCCCAAG 914

#### RESULT 13

US-09-923-556-1/c  
Sequence 1, Application US/09923556  
Patent No. US20020058321A1  
GENERAL INFORMATION:  
APPLICANT: Busch, Marco  
APPLICANT: Hain, Rudiger  
APPLICANT: Martin, William  
APPLICANT: Tietjen, Klaus  
APPLICANT: Kloti, Andreas  
TITLE OF INVENTION: Method of determining the activity of  
TITLE OF INVENTION: 1-deoxy-D-xylulose-5-phosphate reductoisomerase and  
TITLE OF INVENTION: 1-deoxy-D-xylulose-5-phosphate synthase  
FILE REFERENCE: 2020US  
CURRENT APPLICATION NUMBER: US/09/923,556  
CURRENT FILING DATE: 2001-08-07  
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/449,335  
PRIOR FILING DATE: EARLIER FILING DATE: 1999-11-24  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 1732  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (18)..(1448)  
OTHER INFORMATION: DNA encoding 1-deoxy-D-xylulose-5-phosphate  
reductoisomerase  
US-09-923-556-1

Query Match

Best Local Similarity 100.0%; Pred. No. 10;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 195 TTGGATTACGCTGGAGA 211  
|||||

Db 55 TTGGATTACGCTGGAGA 39

#### RESULT 14

US-09-923-556-5/c  
Sequence 5, Application US/09923556  
Patent No. US20020058321A1  
GENERAL INFORMATION:  
APPLICANT: Busch, Marco  
APPLICANT: Hain, Rudiger  
APPLICANT: Martin, William  
APPLICANT: Tietjen, Klaus  
APPLICANT: Kloti, Andreas  
TITLE OF INVENTION: Method of determining the activity of  
TITLE OF INVENTION: 1-deoxy-D-xylulose-5-phosphate reductoisomerase and  
TITLE OF INVENTION: 1-deoxy-D-xylulose-5-phosphate synthase  
FILE REFERENCE: 2020US  
CURRENT APPLICATION NUMBER: US/09/923,556  
CURRENT FILING DATE: 2001-08-07  
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/449,335  
PRIOR FILING DATE: EARLIER FILING DATE: 1999-11-24  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 5  
LENGTH: 1732  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (18)..(1448)  
OTHER INFORMATION: new coding sequence  
US-09-923-556-5

Query Match 4.8%; Score 17; DB 10; Length 1732;

Best Local Similarity 100.0%; Pred. No. 10;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 195 TTGGATTACGCTGGAGA 211  
|||||

Db 55 TTGGATTACGCTGGAGA 39

#### RESULT 15

US-09-764-864-250/c  
Sequence 250, Application US/09764864  
Patent No. US20020132753A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PT223  
CURRENT APPLICATION NUMBER: US/09/764,864  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 1792  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 250  
LENGTH: 2131  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-764-864-250

Query Match

Best Local Similarity 100.0%; Pred. No. 10;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 TCTGTCTCTCCAGTTTTC 116  
|||||

Db 90 TCTGTCTCTCCAGTTTTC 74

Search completed: November 6, 2002, 19:26:17  
Job time : 186 secs

---

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: November 6, 2002, 15:28:31 ; Search time 2766 Seconds  
(without alignments)  
3693.089 Million cell updates/sec  
Title: US-09-897-438b-1  
Perfect score: 351  
Sequence: 1 gagcagtgtggcaccatcat.....atgagcctgtgtggccctg 351  
Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0  
Searched: 2054640 seqs, 14551402878 residues  
Word size : 0  
Total number of hits satisfying chosen parameters: 4109280  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Listing first 45 summaries

Database :

1:	gb.ba.*	GenEmbl.*
2:	gb.htg.*	
3:	gb.in.*	
4:	gb.om.*	
5:	gb.ov.*	
6:	gb.pat.*	
7:	gb.ph.*	
8:	gb.pl.*	
9:	gb.pr.*	
10:	gb.ro.*	
11:	gb.sts.*	
12:	gb.sy.*	
13:	gb.un.*	
14:	gb.vi.*	
15:	em.ba.*	
16:	em.fun.*	
17:	em.hum.*	
18:	em.in.*	
19:	em.mu.*	
20:	em.om.*	
21:	em.or.*	
22:	em.ov.*	
23:	em.pat.*	
24:	em.ph.*	
25:	em.pl.*	
26:	em.ro.*	
27:	em.sts.*	
28:	em.un.*	
29:	em.vi.*	
30:	em.htg_hum.*	
31:	em.htg_inv.*	
32:	em.htg_other.*	
33:	em.htg_mus.*	
34:	em.htg_pln.*	
35:	em.htg_rod.*	
36:	em.htg_mam.*	
37:	em.htg_vrt.*	
38:	em_sy.*	
39:	em_htgo_hum.*	
40:	em_htgo_mus.*	
41:	em_htgo_other.*	

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	
351	100.0	11673	6	AX305377	Sequence		
351	100.0	11673	10	MMU24703	U24703 Mus musculus		
351	135	38.5	163985	10	AC121878	AC121878 Mus muscu	★
4	135	38.5	183156	2	AC023062	AC023062 Mus muscu	
5	33	9.4	126130	2	AC095877	AC095877 Rattus no	
6	33	9.4	183641	2	AC124933	AC124933 Rattus no	
7	33	9.4	202764	2	AC128022	AC128022 Rattus no	
8	32	9.1	11187	10	AB049473	AB049473 Rattus no	
9	26	7.4	11580	6	AX410790	AX410790 Sequence	
10	26	7.4	11580	9	HSU79716	U79716 Human reelin	
11	26	7.4	11580	11	G30936	G30936 swss2926 Er	
12	26	7.4	11580	11	G30938	G30938 swss3176 Er	
13	24	6.8	70905	2	AC119906	AC119906 Mus muscu	
14	21	6.0	73070	9	AL590143	AL590143 Human DNA	
15	21	6.0	98873	2	AC015229	AC015229 Drosophil	
16	21	6.0	160883	2	AL158814	AL158814 Homo sapi	
17	21	6.0	168652	3	AC009384	AC009384 Drosophil	
18	21	6.0	169059	2	AC009790	AC009790 Homo sapi	
19	21	6.0	192491	2	AC103705	AC103705 Homo sapi	
20	21	6.0	298386	3	AE003514	AE003514 Drosophil	
21	20	5.7	6310	5	ACSWSPS3	AF134194 Anolis ca	
22	20	5.7	46890	2	AC109654	AC109654 Rattus no	
23	20	5.7	128913	2	AF004053	AF004053 Oryza sat	
24	20	5.7	164192	9	AC074338	AC074338 Human Chr	
25	20	5.7	176163	2	AC099604	AC099604 Mus muscu	
26	20	5.7	185652	9	AC007312	AC007312 Homo sapi	
27	20	5.7	208405	2	AC079477	AC079477 Mus muscu	
28	19	5.4	937	9	HSDBHGT	X63418 H.sapiens (	
29	19	5.4	2796	9	AK095398	AK095398 Homo sapi	
30	19	5.4	3452	9	AK091510	AK091510 Homo sapi	
31	19	5.4	4883	6	BD012821	BD012821 Y568, a n	
32	19	5.4	4883	23	BD008256	BD008256 Y568, a n	
33	19	5.4	5092	9	HSMB04804	AL833491 Homo sapi	
34	19	5.4	6931	9	HSMB03520	AL832213 Homo sapi	
35	19	5.4	7215	6	BD012826	BD012826 Y568, a n	
36	19	5.4	7215	6	BD012827	BD012827 Y568, a n	
37	19	5.4	7215	9	AB059277	AB059277 Homo sapi	
38	19	5.4	7215	23	BD008261	BD008261 Y568, a n	
39	19	5.4	7215	23	BD008262	BD008262 Y568, a n	
40	19	5.4	37584	9	AC001227	AC001227 Genomic s	
41	19	5.4	44089	9	AC002101	AC002101 Genomic s	
42	19	5.4	61676	2	AC060800	AC060800 Homo sapi	
43	19	5.4	101555	2	AC095962	AC095962 Rattus no	
44	19	5.4	144581	2	AC124868	AC124868 Rattus no	
45	19	5.4	157266	2	AC127765	AC127765 Rattus no	

ALIGNMENTS

RESULT 1	AX305377	11673 bp	DNA	linear	PAT 11-DEC-2001
LOCUS	Sequence 128 from Patent WO0188188.				
DEFINITION	AX305377				
ACCESSION	AX305377				
VERSION	AX305377.1	GI:17644926			
KEYWORDS	house mouse.				
SOURCE	Mus musculus				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
REFERENCE	1				
AUTHORS	Ishikawa, K., Asai, S., Takahashi, Y., Nagata, T. and Ishii, Y.				
TITLE	Method for examining ischemic conditions				
JOURNAL	Patent: WO 0188188-A 128 22-NOV-2001;				

School Juridical Person Nihon University (JP)

## FEATURES

Location/Qualifiers  
1. 11673  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
BASE COUNT 2831 a 2985 c 2985 g 2872 t  
ORIGIN

5'UTR  
CDS

Query Match 100.0%; Score 351; DB 6; Length 11673;

Best Local Similarity 100.0%; Pred. No. 9.7e-193;  
Matches 351; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAGCAGTGTGCGACCATCATGCGAATGCTCTCACCCTTCTGTGAGCCGCTACGGCCCT 60  
Db 970 GAGCAGTGTGCGACCATCATGCGAATGCTCTCACCCTTCTGTGAGCCGCTACGGCCCT 1029  
Qy 61 CGAGAGCTGACCAACACATGCTCAACACACACACACATGCTCCCTCCAGTTTCCATT 120  
Db 1030 CGAGAGCTGACCAACACATGCTCAACACACACACACATGCTCCCTCCAGTTTCCATT 1089  
Qy 121 GGGTCAGGATCATGCTGATTTAGTTTACTCTGACCCAGCATCACTGTCTCATACGCCAAG 180  
Db 1090 GGGTCAGGATCATGCTGATTTAGTTTACTCTGACCCAGCATCACTGTCTCATACGCCAAG 1149  
Qy 181 AACAAATACCGTGTGATTCAGCTGAGGAAATTTAGAGCCCTTCCAAATGTGAGCACA 240  
Db 1150 AACAAATACCGTGTGATTCAGCTGAGGAAATTTAGAGCCCTTCCAAATGTGAGCACA 1209  
Qy 241 GTCATCCACATCCTGTACCTCCCGAGGAGCAAGCAAGGAGAGCGTCCAGTCCAGTGG 300  
Db 1210 GTCATCCACATCCTGTACCTCCCGAGGAGCAAGCAAGGAGAGCGTCCAGTCCAGTGG 1269  
Qy 301 AAACAGGACAGCCTGCGAGTGGGTGAGTGTATGAGGCGCTGCTGGGCCCTG 351  
Db 1270 AAACAGGACAGCCTGCGAGTGGGTGAGTGTATGAGGCGCTGCTGGGCCCTG 1320

## RESULT 2

MMU24703 11673 bp mRNA linear ROD 19-DEC-1997  
LOCUS  
DEFINITION Mus musculus reelin mRNA, complete cds.  
ACCESSION U24703  
VERSION U24703.1 GI:2702252  
KEYWORDS  
SOURCE mouse.

## ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 11673)  
D'Arcangelo, G., Miao, G.G., Chen, S.C., Soares, H.D., Morgan, J.I. and Curran, T.

A protein related to extracellular matrix proteins deleted in the mouse mutant reeler  
Nature 374 (6524), 719-723 (1995)  
95231649

## MEDLINE

PUBMED 7715726

2- (bases 1 to 11673)

D'Arcangelo, G.

Direct Submission

Submitted (10-APR-1995) Roche Institute of Molecular Biology, 340  
Kingsland St., Nutley, NJ 07110, USA  
3 (bases 1 to 11673)  
D'Arcangelo, G.

Direct Submission

Submitted (19-DEC-1997) Dev. Neurobiology, St. Jude Childr. Res.

Hosp. 332 N. Lauderdale, Memphis, TN 38105, USA

Sequence update by submitter

On Dec 19, 1997 this sequence version replaced gi:902486.

## REMARK

Location/Qualifiers

1. 11673

/organism="Mus musculus"

/db\_xref="taxon:10090"

/chromosome="5"

## FEATURES

3'UTR

polyA\_signal

11662..11667

BASE COUNT 2831 a 2985 c 2985 g 2872 t

ORIGIN

/map="D5Gmr1"  
/tissue\_type="cerebellum"  
/dev\_stage="adult"  
1. 282  
283..10668  
/codon\_start=1  
/product="reelin"  
/protein\_id="AAB91599.1"  
/db\_xref="GI:2702253"  
/translation="MERGCGAPRALVLAIVLLLATLRARATGYPRFSPFFFLCTHH  
GELEGEQEVLSIHAGNPYYVPGQEHVHTISTSTPFDGLLVGLTSTSIQSS  
QSIGGSAGFQIMSDHQFQMGCMCVASHVSHPTNLNFWIAPPPAGTGVNFMA  
TATRHQVIFKDALAQCLCEQATATAYSHLAEIHSVILRDVDSVQQLNPNP  
IWECSNCEMGKQGTIMHGNATVFCPEPNVRLTCLNTTASVLSIGSGCRF  
SYDPSITVYAKNNTADTQLEKIRAPSNVTVIHLYPEAKQESVQFQWQDSL  
RGEVVEACWALDNILVINSAREVLEONLDPVDTGNMLFFPGATVKKSCQDGSNI  
YFNGEGSENFATRDVDLSTEDIQEQSEEFESQPTGWDILGACVACDGVESGL  
SLVFLKDGKCKICTPYMDTGTGNLRFYVMGICDGPVSHENIILYAKIERKEHI  
ADLTLYSSYKPSLVSVINLQTPATKCLRQKSHOGYNRNWAVDFHVLPLP  
STSMHIOFSINLGCTHOPGNSVLEFSTNHRGWSLLHTECLPETCAGPLHPSIV  
YSENTSGNRIITPLNAAITRDTRIRWOTGPILGNMAIDNVYISGAEVFCGR  
GOCTRHGCKDCPGFCPACEMASOTPEMTISESFGSARLSYHNFYISGAEVFCGR  
VLASGALVFNKDRQLTSLFSDSSQSRPLQFTLRGSKSVLSTCRAPDQPEGVLL  
HYSYDNGITWKLLEHSYVNIHEPIISVDELDPDAROFQIFQFRWQPYHSSQSDVMA  
IDIVMTSVLNSISLDTNLNVEVTSGLFYLGNOVQYCGHDWTLCTGDKSLASMA  
YVETQSMQIGASYMIOFSLVMGCGQYTPHMDNQVKLEYSAHGLTWHLVQEECLPSM  
PSCQETASISIHASEFTWRVTVYLPKOTWSGATFRWSQSYTAQDEWALDNTVI  
GOCPNMCSHGSDHGVCRDQYOGTECHPEALPSTIMSPDPSWESDQSVI  
GGVVEPEQCGVYSSGSSLYSKAGKRLVSHDLTSDWDFVQVFIQGGESAACNK  
PDSREGILLIYNNGGIOHLLIAEMTFSDFSKPRFVYELPAAGTKPTFRWPKPV  
FSGEDYQWAVADDIIILSEKQKQIPVNPPTLPQNFYEKPAFDYPNMOKSVMLANE  
GMAKNSFCATPSAMVFGKSDGDFAVTRDLTKPGYVLOQLNGICTSQFSTAPV  
LLQYSHDAGMSWFLKGGCFPPASAAKCGNSRELSEPTVYTGDFEEMTRITAIAPR  
SLASKTRFRIOEISSOKNVPPEGLDGVYISPCPSYCGHGDCISGVCFDLGYTA  
VRYVPLVPLTPRFRWIOYNTVGADSWAIDNVYILASGCPMWCSCGICDSCRC  
VCDRGFGGCPVVPVPLPSLTKDQFNGLHPLWDMVYGAERGNLNGETLKSGCTLI  
KGEGLMLISRLDCLNTMYVQPSLRFIAKGTPEPERSHSLILQFSVSGGVTHLMDREY  
FPOTTSILFINVLPYGAQTNRFRWQPYNNGKKEEIIWIDFIIDGNLNNPVLL  
LDTFDFGREDNMFYGGNIGLYCPSKGAPEEDSAMVSEVGEHSITTRDLSV  
NENTIIQFEINVCSDSSADPVRLEFSDFGATWHLLPLCYHSSLSVSLCSTE  
HPSYTAGTTCGWRREVHFGKLHCGSVFRWYOGFYAGSQPYTWAIDNVYICGP  
CREMCGHSCINGTACICDGVSGPCKISTKNPDLKDDPEGQLESOREFLMSGK  
PRKCGILSGNNLNFENEDGLRMLVTRDLDSHARVQFFMRIGLCKGKVPDRSQPVL  
LQYSLNGSWSLLEFLEFSNSNVGRYIALEMLKARSGSTRLRWQPSENGHFYSP  
WYIDQLIGNISGNTVLEDFSTLDSRWKLLHPGCTKMPVCGSTGALVFEKASTR  
YVYVTDIAVNEDSFLQIDFAACSVTDSCVATELEYSDVLGSLWHLVRAIDNVYICG  
GYHLQRIIVSDTFNKWTRITLPLPSYTSQATRRFRWOPADPKOOTWADNVYIGD  
SCLDMCSGHRGCVQSCVDEQWGLYCDPEPSTLQKLAKNFNAPSQNWQVSGG  
KLSTVCGAVAGLALHFSGCKSLVLDNLNTNAEFIQIFYMIGCLITPSNRQWGLV  
LEYSVNGITWNLLEIFDYQSKPGFVNLPLPDAKEATATFRWOPRHDGDLQNDW  
AIDNVLISGADQRTVMDTFSSAPVQHERSPADAPGVGFATFLEBDKTSVNEKN  
LFPHDCTVERFCSDPGVMCLSGHSDGREVAVTHDLTPENIMQKFSVCGKVPKPI  
AQNOIHQVSTDFGVSWSYLVQCLPADPKCSGSVSQSPVFPTEGKRTTYPLPSL  
TGNVPFRFYQKYSYDQWADNFYLGPGCLDNGCGHGDCLKEQICDQPSGPNCLYT  
HSLKTLKRFDESEIKPDLMSLEGSTCTEGVLAENTALFYFGSTVQRAITQDLD  
LRGAKFLQYWRGISENNMTSCHRPVREKGLLDFTSDGTGITWLLHMDFOKYSV  
RHDYILLPEGALTNRRLWQPVFISNGLVSVGVAQWALDNILGAEINPSOLV  
DTPDDEGSSHEENWSPYNAVATAGCPNPSFLWPNKKKDKTHALSSRELIIQOP  
YMWQKIVVCEATSCGDLHSMVLETKDARSDMQLVOTQCLPSSNSISGSPGFH  
EATINAVNSSSKRTIOLPHVSSSATQFRWIOKETEKEQSWAIDNVYIGECAPK  
LCSGHGYCTTGAVICDESDGDCSVFHELPSYIKDNFESARVATWETIQGGVI  
SGCGQLAPAHGDSLYFNGCQIRQAATPLDLTRKATFVLIQIGSPQSDCL  
SGPHTVAKVLLQYSVNNCTIHWIAHQPKDFTQARVSYNVPLEARKMGLRWQ  
PRHNGTGHQWALDHVEVVLVSTRKQNYMNFESRQHLRHFYNNRRRSRLRPV"  
10669..11673

```
Query Match      100.0%; Score 351; DB 10; Length 11673;
Best Local Similarity 100.0%; Pred. No. 9.7e-193;
Matches 351; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGCAGTGTGGCACCATCATGCGAATGCTGTACCTTCTGTGAGCCGTAGCGCCCT 60
    |||||||
Db 970 GAGCAGTGTGGCACCATCATGCGAATGCTGTACCTTCTGTGAGCCGTAGCGCCCT 1029
    |||||||

QY 61 CGAGAGTGCACCAACCATCCCTGAACACAAACAGCATCTGTGCTCCAGTTTCCATT 120
    |||||||
Db 1030 CGAGAGTGCACCAACCATCCCTGAACACAAACAGCATCTGTGCTCCAGTTTCCATT 1089
    |||||||

QY 121 GGTCAGGATCATGTCGATTTAGTTACTCTGACCCAGCATCTGTGTATATAGCCCAAG 180
    |||||||
Db 1090 GGTCAGGATCATGTCGATTTAGTTACTCTGACCCAGCATCTGTGTATATAGCCCAAG 1149
    |||||||

QY 181 AACAAATACCCGCTGATTCGATTCAGTCGAGAAATAGAGCCCTTCCAAATGTGAGCACA 240
    |||||||
Db 1150 AACAAATACCCGCTGATTCGATTCAGTCGAGAAATAGAGCCCTTCCAAATGTGAGCACA 1209
    |||||||

QY 241 GTCATCCACATCTGTACCTCCCGAGGAGCAAGCCGAGGAGCGTGCAGTTCAGTGG 300
    |||||||
Db 1210 GTCATCCACATCTGTACCTCCCGAGGAGCAAGCCGAGGAGCGTGCAGTTCAGTGG 1269
    |||||||

QY 301 AAACAGACAGCCCTGCAGTGGGTGAGGTGTATGAGCCCTGCTGGGCCCTG 351
    |||||||
Db 1270 AAACAGACAGCCCTGCAGTGGGTGAGGTGTATGAGCCCTGCTGGGCCCTG 1320
    |||||||

RESULT 3
AC121878/c
LOCUS      AC121878      163985 bp      DNA      linear      ROD 21-JUN-2002
DEFINITION Mus musculus clone RP24-131C8, complete sequence.
ACCESSION  AC121878
VERSION    AC121878.1 GI:21039993
KEYWORDS   HTG.
SOURCE     house mouse.
ORGANISM   Mus musculus
REFERENCE  1 (bases 1 to 163985)
AUTHORS    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
TITLE      The sequence of Mus musculus clone
JOURNAL    Unpublished
AUTHORS     McPherson, J.D. and Waterston, R.H.
JOURNAL    Direct Submission
TITLE      Submitted (21-MAY-2002) Genome Sequencing Center, 4444 Forest Park
JOURNAL    Parkway, St. Louis, MO 63108, USA
REFERENCE  3 (bases 1 to 163985)
AUTHORS     McPherson, J.D. and Waterston, R.H.
JOURNAL    Direct Submission
TITLE      Submitted (21-JUN-2002) Genome Sequencing Center, 4444 Forest Park
JOURNAL    Parkway, St. Louis, MO 63108, USA
COMMENT

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: M_BB0131C08
----- Location/Qualifiers -----
1. .163985
   /organism="Mus musculus"
   /db_xref="taxon:10090"
   /clone="RP24-131C8"
BASE COUNT 50392 a 33960 c 32422 g 47211 t
ORIGIN

Query Match      38.5%; Score 135; DB 10; Length 163985;
```

```
Best Local Similarity 100.0%; Pred. No. 6e-67;
Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 217 AGAGCCCTTCCAAATGTGAGCAGATCATCCACATCTGTACCTCCCGAGAGCCCAA 276
    |||||||
Db 15382 AGAGCCCTTCCAAATGTGAGCAGATCATCCACATCTGTACCTCCCGAGAGCCCAA 15323
    |||||||

QY 277 GGGGAGAGCGTGCAGTTCCAGTGGAAACAGACAGCCCTGCGAGTGGGTGAGGTATGAG 336
    |||||||
Db 15322 GGGGAGAGCGTGCAGTTCCAGTGGAAACAGACAGCCCTGCGAGTGGGTGAGGTATGAG 15263
    |||||||

QY 337 GCCTGCTGGGCCCTG 351
    |||||||
Db 15262 GCCTGCTGGGCCCTG 15248
    |||||||

RESULT 4
AC023062      183156 bp      DNA      linear      HTG 05-MAY-2000
LOCUS      AC023062
DEFINITION Mus musculus chromosome 5 clone RP23-466D2 strain C57BL6/J, WORKING
            DRAFT SEQUENCE, 8 unordered pieces.
ACCESSION  AC023062
VERSION    AC023062.2 GI:7709916
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     Mus musculus.
ORGANISM   Mus musculus
REFERENCE  1 (bases 1 to 183156)
AUTHORS     Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W.,
            Bouffard, G.G., Dietrich, N.L., Eagle, W.O., Gupta, J., Ho, S.-L.,
            Huang, M.C., Idol, J., Lee-Lin, S.-Q., Maduro, Q.L., Maduro, V.B.,
            Mastriano, S.D., McCloskey, J.C., Morse, E., Ojodu, M.A., Pearson, R.,
            Stantripop, S., Summers, T.J., Thomas, J.W., Thomas, P.J.,
            Tingson, E.E., Touchman, J.W., Tran, J.T., Vogt, J.L., Walker, M.A.,
            Wetherby, K.D. and Green, E.D.
JOURNAL    NISC Mouse Sequencing Initiative
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 183156)
AUTHORS     Green, E.D.
TITLE      Direct Submission
JOURNAL    Submitted (08-FEB-2000) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Galthersburg, MD 20877, USA
COMMENT    On May 5, 2000 this sequence version replaced gi:6939141.
            ----- Genome Center -----
            Center: NIH Intramural Sequencing Center
            Center code: NISC
            Web site: http://www.nisc.nih.gov
            Contact: nisc_mouse@nhgri.nih.gov
            ----- Project Information -----
            Center project name: Y1
            Center clone name: 466D02
            ----- Summary Statistics -----
            Sequencing vector: plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 175371 bases at least Q40
            Consensus quality: 178474 bases at least Q30
            Consensus quality: 180133 bases at least Q20
            Insert size: 179000; agarose-fp
            Insert size: 188000; pulse-field-gel
            Quality coverage: 6.96x in Q20 bases; agarose-fp
            Quality coverage: 6.63x in Q20 bases; pulse-field-gel
            Quality coverage: 6.80x in Q20 bases; sum-of-contigs
            -----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
```

\* be preserved.

```

1 2374: contig of 2374 bp in length
* gap of unknown length
* 2375 4490: contig of 2116 bp in length
* gap of unknown length
* 4491 14959: contig of 10469 bp in length
* gap of unknown length
* 14960 31072: contig of 16113 bp in length
* gap of unknown length
* 31073 50119: contig of 19047 bp in length
* gap of unknown length
* 50120 78196: contig of 28077 bp in length
* gap of unknown length
* 78197 109874: contig of 31678 bp in length
* gap of unknown length
* 109875 183156: contig of 73282 bp in length.
  Location/Qualifiers
    1..183156
      /organism="Mus musculus"
      /strain="C57BL6/J"
      /db_xref="taxon:10090"
      /chromosome="5"
      /clone="RP23-466D2"
      /clone_lib="RPCI mouse BAC library 23"
  53719 a 37080 c 37302 g 55052 t 3 others

```

# FEATURES source

BASE COUNT 53719 a 37080 c 37302 g 55052 t 3 others

ORIGIN

```

Query Match      38.5%; Score 135; DB 2; Length 183156;
Best Local Similarity 100.0%; Pred. No. 6e-67;
Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 217 AGAGCCCTCCAAATGTGAGCAGACAGTATCATCTGTACCTGACCTCCGAGGAAGCCAAA 276
      AGAGCCCTCCAAATGTGAGCAGACAGTATCATCTGTACCTGACCTCCGAGGAAGCCAAA 173682

QY 277 GGGGAGACGTCGATCCAGTCCAGTGAACAGCAGCCTGCGAGTGGTGGTGTATGAG 336
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 173683 GGGGAGACGTCGATCCAGTCCAGTGAACAGCAGCCTGCGAGTGGTGGTGTATGAG 173742

QY 337 GCCTGCTGGGCCCTG 351
      |||||||||||||||
Db 173743 GCCTGCTGGGCCCTG 173757

```

RESULT 5

AC095877

LOCUS AC095877 126130 bp DNA linear HTG 11-JUL-2002

DEFINITION Rattus norvegicus clone CH230-10G13, \*\*\* SEQUENCING IN PROGRESS

AC095877 \*\*\*, 59 unordered pieces.

VERSION AC095877.4 GI:21722955

KEYWORDS HTG; HTGS\_PHASE1.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 126130)

AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbara,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhaý,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,

Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C., Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseghe,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,S., Nguyen,N., Nickerson,E., Nwokewo,S., Ogih,M., Okwoudu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaik,E., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.

TITLE JOURNAL

REFERENCE 2 (bases 1 to 126130)

AUTHORS Worley,K.C.

TITLE JOURNAL

REFERENCE 3 (bases 1 to 126130)

AUTHORS Worley,K.C.

TITLE JOURNAL

COMMENT

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GDRS

Center clone name: CH230-10G13

----- Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 70934 bases at least Q40

Consensus quality: 75985 bases at least Q30

Consensus quality: 80731 bases at least Q20

-----

\* NOTE: Estimated insert size may differ from sequence length

\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 59 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 1382: contig of 1382 bp in length

\* 1383 1482: gap of unknown length

\* 1483 3046: contig of 1564 bp in length

\* 3047 3146: gap of unknown length

\* 3147 4489: contig of 1343 bp in length

\* 4490 4590: gap of unknown length

\* 4590 5786: contig of 1197 bp in length

\* 5787 5886: gap of unknown length  
\* 5887 7115: contig of 1229 bp in length  
\* 7116 7115: gap of unknown length  
\* 7216 8771: contig of 1556 bp in length  
\* 8772 8771: gap of unknown length  
\* 8872 10260: contig of 1389 bp in length  
\* 10261 10360: gap of unknown length  
\* 10361 11370: contig of 1010 bp in length  
\* 11371 11470: gap of unknown length  
\* 11471 12710: contig of 1240 bp in length  
\* 12711 12810: gap of unknown length  
\* 12811 14180: contig of 1370 bp in length  
\* 14181 14280: gap of unknown length  
\* 14281 16048: contig of 1768 bp in length  
\* 16049 16148: gap of unknown length  
\* 16149 17569: contig of 1421 bp in length  
\* 17570 17659: gap of unknown length  
\* 17670 18820: contig of 1151 bp in length  
\* 18821 18920: gap of unknown length  
\* 18921 20209: contig of 1289 bp in length  
\* 20210 20309: gap of unknown length  
\* 20310 21633: contig of 1324 bp in length  
\* 21634 21733: gap of unknown length  
\* 21734 23269: contig of 1536 bp in length  
\* 23270 23369: gap of unknown length  
\* 23370 24647: contig of 1278 bp in length  
\* 24648 24747: gap of unknown length  
\* 24748 25800: contig of 1053 bp in length  
\* 25801 27017: contig of 1117 bp in length  
\* 27018 27117: gap of unknown length  
\* 27118 29183: contig of 2066 bp in length  
\* 29184 29283: gap of unknown length  
\* 29284 30315: contig of 1032 bp in length  
\* 30316 32086: contig of 1671 bp in length  
\* 32087 32186: gap of unknown length  
\* 32187 33395: contig of 1209 bp in length  
\* 33396 33495: gap of unknown length  
\* 33496 35663: contig of 2168 bp in length  
\* 35664 35763: gap of unknown length  
\* 35764 38044: contig of 2281 bp in length  
\* 38045 38144: gap of unknown length  
\* 38145 39845: contig of 1701 bp in length  
\* 39846 39945: gap of unknown length  
\* 39946 41982: contig of 2037 bp in length  
\* 41983 42083: gap of unknown length  
\* 42083 43743: contig of 1661 bp in length  
\* 43744 43843: gap of unknown length  
\* 43844 45904: contig of 2061 bp in length  
\* 45905 46004: gap of unknown length  
\* 46005 47667: contig of 1663 bp in length  
\* 47668 47767: gap of unknown length  
\* 47768 49833: contig of 2066 bp in length  
\* 49834 49933: gap of unknown length  
\* 49934 51434: contig of 1501 bp in length  
\* 51435 51534: gap of unknown length  
\* 51535 53778: contig of 2244 bp in length  
\* 53779 53878: gap of unknown length  
\* 53879 55552: contig of 1674 bp in length  
\* 55553 55652: gap of unknown length  
\* 55653 57671: contig of 2018 bp in length  
\* 57671 57770: gap of unknown length  
\* 57771 59630: contig of 1860 bp in length  
\* 59631 59730: gap of unknown length  
\* 59731 61147: contig of 1417 bp in length  
\* 61148 61247: gap of unknown length  
\* 61248 63995: contig of 2748 bp in length  
\* 63996 64095: gap of unknown length  
\* 64095 66294: contig of 2199 bp in length  
\* 66294 66394: gap of unknown length  
\* 66395 68794: contig of 2400 bp in length  
\* 68795 68894: gap of unknown length

\* 68895 71533: contig of 2639 bp in length  
\* 71534 71633: gap of unknown length  
\* 71634 73469: contig of 1836 bp in length  
\* 73470 73569: gap of unknown length  
\* 73570 77054: contig of 3485 bp in length  
\* 77055 77154: gap of unknown length  
\* 77155 79005: contig of 1851 bp in length  
\* 79006 79105: gap of unknown length  
\* 79106 81294: contig of 2189 bp in length  
\* 81295 81394: gap of unknown length  
\* 81395 83864: contig of 2470 bp in length  
\* 83865 83964: gap of unknown length  
\* 83965 86039: contig of 2075 bp in length  
\* 86040 86139: gap of unknown length  
\* 86140 88646: contig of 2507 bp in length  
\* 88647 88746: gap of unknown length  
\* 88747 91823: contig of 3077 bp in length  
\* 91824 91923: gap of unknown length  
\* 91924 94369: contig of 2446 bp in length  
\* 94370 94469: gap of unknown length  
\* 94470 96886: contig of 2417 bp in length  
\* 96887 96986: gap of unknown length  
\* 96987 100680: contig of 3693 bp in length  
\* 100680 100779: gap of unknown length  
\* 100780 104171: contig of 3392 bp in length  
\* 104172 104271: gap of unknown length  
\* 104272 106760: contig of 2489 bp in length

## Query Match

Best Local Similarity 9.4%; Score 33; DB 2: Length 126130;

Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 285 CGTGCAGTTCAGTGGAAACAGACGACCTGCG 317

|||||

Db 92895 CGTGCAGTTCAGTGGAAACAGACGACCTGCG 92927

## RESULT 6

AC124933/c

## LOCUS

Rattus norvegicus clone CH230-466n17, \*\*\* SEQUENCING IN PROGRESS

\*\*\*, 88 unordered pieces.

AC124933.2 GI:21952716

HTG: HTGS\_PHASE1.

Norway rat.

## ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 183641)

## REFERENCE

## AUTHORS

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,  
Barbaria,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,  
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,  
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
Davila,M.L., Davis,C., Davy-Catroll,L., Dederich,D.A.,  
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
Gabis,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,  
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,  
Homsí,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,  
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
Karlssoon,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louiseged,H.,

Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogwu, M., Okwuonu, G., Oragunye, N., Owiedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I., Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tanssey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.	9579	11061: contig of 1483 bp in length
Direct Submission		
Unpublished		
2 (bases 1 to 183641)		
Worley, K.C.		
Direct Submission		
Submitted (20-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA		
3 (bases 1 to 183641)		
Worley, K.C.		
Direct Submission		
Submitted (31-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA		
On Jul 24, 2002 this sequence version replaced gi:21490053.		
----- Genome Center		
Center: Baylor College of Medicine		
Center code: BCM		
Web site: <a href="http://www.hgsc.bcm.tmc.edu/">http://www.hgsc.bcm.tmc.edu/</a>		
Contact: hgsc-help@bcm.tmc.edu		
----- Project Information		
Center project name: KAVI		
Center clone name: CH230-466N17		
----- Summary Statistics		
Sequencing vector: Plasmid;		
Chemistry: Dye-terminator Big Dye; 100% of reads		
Assembly program: Phrap; version 0.990329		
Consensus quality: 111724 bases at least Q40		
Consensus quality: 120970 bases at least Q30		
Consensus quality: 128458 bases at least Q20		
-----		
* NOTE: Estimated insert size may differ from sequence length (see <a href="http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html">http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html</a> ).		
* NOTE: This is a 'working draft' sequence. It currently consists of 88 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.		
* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.		
1 1107: contig of 1107 bp in length		
* 1108 1207: gap of unknown length		
* 1208 2460: contig of 1253 bp in length		
* 2461 2560: gap of unknown length		
* 2561 4004: contig of 1444 bp in length		
* 4005 4104: gap of unknown length		
* 4105 5382: contig of 1278 bp in length		
* 5383 5482: gap of unknown length		
* 5483 6874: contig of 1392 bp in length		
* 6875 6974: gap of unknown length		
* 6975 8071: contig of 1097 bp in length		
* 8072 8171: gap of unknown length		
* 8172 9478: contig of 1307 bp in length		
* 9479 9578: gap of unknown length		
* 9579 11061: contig of 1483 bp in length		
* 11062 11161: gap of unknown length		
* 11162 12742: contig of 1581 bp in length		
* 12743 12842: gap of unknown length		
* 12843 14504: contig of 1662 bp in length		
* 14505 14604: gap of unknown length		
* 14605 15734: contig of 1130 bp in length		
* 15735 15834: gap of unknown length		
* 15835 16958: contig of 1124 bp in length		
* 16959 17058: gap of unknown length		
* 17059 18704: contig of 1646 bp in length		
* 18705 18804: gap of unknown length		
* 18805 20420: contig of 1616 bp in length		
* 20421 20520: gap of unknown length		
* 20521 21648: contig of 1128 bp in length		
* 21649 21748: gap of unknown length		
* 21749 23177: contig of 1429 bp in length		
* 23178 23277: gap of unknown length		
* 23278 24760: contig of 1483 bp in length		
* 24761 24860: gap of unknown length		
* 24861 26759: contig of 1899 bp in length		
* 26760 26859: gap of unknown length		
* 26860 28299: contig of 1440 bp in length		
* 28300 28399: gap of unknown length		
* 28400 29592: contig of 1193 bp in length		
* 29593 29692: gap of unknown length		
* 29693 30870: contig of 1178 bp in length		
* 30871 30970: gap of unknown length		
* 30971 32642: contig of 1672 bp in length		
* 32643 32742: gap of unknown length		
* 32743 34386: contig of 1644 bp in length		
* 34387 34486: gap of unknown length		
* 34487 36064: contig of 1578 bp in length		
* 36065 36164: gap of unknown length		
* 36165 37398: contig of 1234 bp in length		
* 37399 37498: gap of unknown length		
* 37499 39608: contig of 2110 bp in length		
* 39609 39708: gap of unknown length		
* 39709 41480: contig of 1772 bp in length		
* 41481 41580: gap of unknown length		
* 41581 42759: contig of 1179 bp in length		
* 42760 42859: gap of unknown length		
* 42860 44711: contig of 1852 bp in length		
* 44712 44811: gap of unknown length		
* 44812 46208: contig of 1397 bp in length		
* 46209 46308: gap of unknown length		
* 46309 47767: contig of 1459 bp in length		
* 47768 47867: gap of unknown length		
* 47868 49227: contig of 1360 bp in length		
* 49228 49327: gap of unknown length		
* 49328 50797: contig of 1470 bp in length</		



```

* 66084 66183: gap of unknown length
* 66184 67516: contig of 1333 bp in length
* 67517 67616: gap of unknown length
* 67617 69820: contig of 2204 bp in length
* 69821 69920: gap of unknown length
* 69921 71261: contig of 1341 bp in length
* 71262 71361: gap of unknown length
* 71362 72775: contig of 1414 bp in length
* 72776 72875: gap of unknown length
* 72876 75298: contig of 2423 bp in length
* 75299 75398: gap of unknown length
* 75399 77094: contig of 1696 bp in length
* 77095 77194: gap of unknown length
* 77195 78834: contig of 1640 bp in length
* 78835 78934: gap of unknown length
* 78935 79989: contig of 1055 bp in length
* 79990 80089: gap of unknown length
* 80090 81774: contig of 1685 bp in length
* 81775 81874: gap of unknown length
* 81875 83255: contig of 1381 bp in length

Query Match          9.4%; Score 33; DB 2; Length 183641;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 285 CGTGCAGTTCCTGAGTGAACAGACAGACGCTGCG 317
Db 143654 CGTGCAGTTCCTGAGTGAACAGACAGACGCTGCG 143622
|||||
CGTGCAGTTCCTGAGTGAACAGACAGACGCTGCG 317
CGTGCAGTTCCTGAGTGAACAGACAGACGCTGCG 143622

RESULT 7
AC128022/c
LOCUS      AC128022
DEFINITION Rattus norvegicus clone CH230-525L20, *** SEQUENCING IN PROGRESS
ACCESSION AC128022
VERSION    AC128022.1 GI:21908605
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE     Rattus norvegicus.
ORGANISM   Rattus norvegicus.
REFERENCE  1 (bases 1 to 202764)
AUTHORS    Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-oshman, F.R., Allen, C.,
            Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Ayele, M., Banks, T.,
            Barbarella, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D.,
            Bouckle, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
            Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
            Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
            Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
            Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
            Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
            Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
            Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
            Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
            Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
            Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
            Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
            Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
            Hernandez, O., Hodgson, A., Hognes, M., Holloway, C., Hollins, B.,
            Homsfi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
            Jacobson, B., Jia, Y., Johnson, R., Jollivet, S., Joudah, S.,
            Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
            Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
            Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H.,
            Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
            Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
            Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
            Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
            Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
            Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G.,
            Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
            Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
            Rives, M., Rojas, A., Rojebokan, I., Rolfe, M., Ruiz, S., Savary, G.,
            Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I.,
            Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,
            Sutton, A., Svatek, A., Tabot, P., Tamerisa, A., Tamerisa, K., Tang, H.,
            Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
            Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
            Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
            Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K.,
            Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
            Weinstock, G. and Gibbs, R.
            Unpublished
            Direct Submission
            2 (bases 1 to 202764)
            Worley, K.C.
            Direct Submission
            Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
            ----- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            ----- Project Information
            Center project name: KAIV
            Center clone name: CH230-525L20
            ----- Summary Statistics
            Sequencing vector: Plasmid;
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 113069 bases at least Q40
            Consensus quality: 119970 bases at least Q30
            Consensus quality: 125656 bases at least Q20
            -----
            * NOTE: Estimated insert size may differ from sequence length
            * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 97 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            * 1244: contig of 1244 bp in length
            * 1245 1344: gap of unknown length
            * 1345 2346: contig of 1002 bp in length
            * 2347 2446: gap of unknown length
            * 2447 3813: contig of 1367 bp in length
            * 3814 3913: gap of unknown length
            * 3914 5201: contig of 1188 bp in length
            * 5102 5201: gap of unknown length
            * 5202 6724: contig of 1523 bp in length
            * 6725 6824: gap of unknown length
            * 6825 8048: contig of 1224 bp in length
            * 8049 8148: gap of unknown length
            * 8149 9212: contig of 1064 bp in length
            * 9213 9312: gap of unknown length
            * 9313 10432: contig of 1120 bp in length
            * 10433 10532: gap of unknown length
            * 10533 11890: contig of 1358 bp in length
            * 11891 11990: gap of unknown length
            * 11991 13068: contig of 1078 bp in length
            * 13069 13169: gap of unknown length
            * 13169 14486: contig of 1318 bp in length
            * 14487 14587: gap of unknown length
            * 14588 15996: contig of 1410 bp in length
            * 15997 16096: gap of unknown length
            * 16097 17534: contig of 1438 bp in length
            * 17535 17634: gap of unknown length
            * 17635 18703: contig of 1069 bp in length
            * 18704 18803: gap of unknown length

```

```

* 18804 19884: contig of 1081 bp in length
* 19885 19885: gap of unknown length
* 19885 21539: contig of 1555 bp in length
* 21540 21639: gap of unknown length
* 21540 22780: contig of 1141 bp in length
* 22781 22880: gap of unknown length
* 22881 23926: contig of 1046 bp in length
* 23927 24026: gap of unknown length
* 24027 25544: contig of 1518 bp in length
* 25545 25644: gap of unknown length
* 25645 26899: contig of 1255 bp in length
* 26900 26999: gap of unknown length
* 27000 28237: contig of 1238 bp in length
* 28238 28337: gap of unknown length
* 28338 29758: contig of 1421 bp in length
* 29759 29858: gap of unknown length
* 29859 31038: contig of 1180 bp in length
* 31039 31138: gap of unknown length
* 31139 32279: contig of 1041 bp in length
* 32180 32279: gap of unknown length
* 32280 33796: contig of 1517 bp in length
* 33797 33896: gap of unknown length
* 33897 35380: contig of 1484 bp in length
* 35381 35480: gap of unknown length
* 35481 37032: contig of 1552 bp in length
* 37033 37132: gap of unknown length
* 37133 38414: contig of 1282 bp in length
* 38415 38514: gap of unknown length
* 38515 39948: contig of 1434 bp in length
* 39949 40048: gap of unknown length
* 40049 41436: contig of 1388 bp in length
* 41437 41536: gap of unknown length
* 41537 43084: contig of 1548 bp in length
* 43085 43184: gap of unknown length
* 43185 44940: contig of 1756 bp in length
* 44941 45040: gap of unknown length
* 45041 46186: contig of 1146 bp in length
* 46187 46286: gap of unknown length
* 46287 47989: contig of 1703 bp in length
* 47990 48089: gap of unknown length
* 48090 49842: contig of 1753 bp in length
* 49843 49942: gap of unknown length
* 49943 51128: contig of 1186 bp in length
* 51129 51228: gap of unknown length
* 51229 52923: contig of 1595 bp in length
* 52924 52923: gap of unknown length
* 52924 54431: contig of 1508 bp in length
* 54432 54531: gap of unknown length
* 54532 56042: contig of 1511 bp in length
* 56043 56142: gap of unknown length
* 56143 57287: contig of 1145 bp in length
* 57288 57387: gap of unknown length
* 57388 58993: contig of 1506 bp in length
* 58994 58993: gap of unknown length
* 58994 60024: contig of 1031 bp in length
* 60025 60124: gap of unknown length
* 60125 62858: contig of 2734 bp in length
* 62859 64365: contig of 1407 bp in length
* 64366 64465: gap of unknown length
* 64466 66434: contig of 1969 bp in length
* 66435 66534: gap of unknown length
* 66535 68632: contig of 2097 bp in length
* 68633 68731: gap of unknown length
* 68732 70839: contig of 2108 bp in length
* 70840 70939: gap of unknown length
* 70940 73009: contig of 2070 bp in length
* 73010 73109: gap of unknown length
* 73110 74957: contig of 1848 bp in length
* 74958 75057: gap of unknown length
* 75058 76816: contig of 1759 bp in length
* 76817 76916: gap of unknown length
* 76917 78679: contig of 1763 bp in length

```

```

* 78680 78779: gap of unknown length
* 78780 80056: contig of 1277 bp in length
* 80057 80156: gap of unknown length
* 80157 81312: contig of 1156 bp in length
* 81313 81412: gap of unknown length
* 81413 82922: contig of 1510 bp in length
* 82923 83022: gap of unknown length
* 83023 84180: contig of 1158 bp in length
* 84181 84280: gap of unknown length
* 84281 85849: contig of 1569 bp in length
* 85850 85949: gap of unknown length
* 85950 87612: contig of 1663 bp in length
* 87613 87712: gap of unknown length

Query Match          9.4%; Score 33; DB 2; Length 202764;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 285 CGTGCAGTTCCAGTGGAAACAGGACAGCCTGCG 317
      |||||
Db 187247 CGTGCAGTTCCAGTGGAAACAGGACAGCCTGCG 187215

RESULT 8
LOCUS      AB049473
DEFINITION Rattus norvegicus mRNA for reelin, complete cds.
ACCESSION AB049473
VERSION    AB049473.1 GI:17221617
KEYWORDS   .
SOURCE      Rattus norvegicus juvenile cerebellum cDNA to mRNA.
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1
AUTHORS    Kikkawa,S. and Terashima,T.
TITLE      rat reelin (Reln) complete CDS
JOURNAL    Published Only in Database (2001)
REFERENCE   2 (bases 1 to 11187)
AUTHORS    Kikkawa,S. and Terashima,T.
TITLE      Direct Submission
JOURNAL    Submitted (29-SEP-2000) Satoshi Kikkawa, Kobe University School of
            Medicine, Department of Anatomy; Chuo-ku, Kusunoki-cho 7-5-1, Kobe,
            Hyogo 650-0017, Japan (E-mail:skikkawa@med.kobe-u.ac.jp,
            Tel:81-78-382-5325; Fax:81-78-382-5328)

FEATURES             Location/Qualifiers
     source            1..11187
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /tissue_type="cerebellum"
                     /dev_stage="juvenile"
     gene              1..11187
                     /gene="Reln"
     CDS                341..10729
                     /gene="Reln"
                     /codon_start=1
                     /product="reelin"
                     /protein_id="BAB78470.1"
                     /db_xref="GI:17221618"
     translation="MERGWAQPTLVLAIVLALLLLATLRARATGYPRSPFFFLCTH
HCELEGDGEQEVGLISLHIAQNTYVPGQYHVITSTFFDGLLVLTGLTSTIQS
SQSIGGSAGFGIMSDHGFQNGFMCVSVASHLPTNLSFVVIAPAGTCGVNFM
ATATHRGQVIFKDALAQCLCEQGAPEATAYSHLAEIHSDSVILRRDDFDSYHQLBNR
NIWACSNCDTGEQCGTTHMGNAVTECPYGPRLTITTYLNTTASVLFQSTGSGSCR
FSYSDPSLIIVYAKNNTADWQLEKIRAPSNVSTIIHILYLPEDAKGENVQPMQRQDS
LHVGEVYEAQWALDNILVNSAHQVILEDSLPDVTGNLFFPGATVKKHSCQSGNA
IYFHGNEQSLNFATRDVLDLSTIQEQWSEEFESQPTGWDILGAVVSGCQTIESG
LSLVFLKDGKRLKLTGYGNLRFYFAMGTCDDPGDSDHENDVILYAKTEGKKEH
IALDTLSYSSYKVPITLVSNVNLPELQTPATKFCRLQKQKHQGNQNVAVDFHVLPII
PSTSHMIQFINSILGCGTHQPGNSVLSFESINHGKRSWLSLLTECLPEICAGPHLPHST
IYSENYSGNWRVTIPLNAAALTRDTRIRWRQTGPIGNMMAIDNVYIGPSCIKFCFSG
RGQCTRHGCKDPCGFSGPACEMASQTFPMFISESGSSRLSSYHNFIYIRGAEVSGFC

```

GVLASGKALVFNKDGRODLITSLFDSSQSRRFLOFTLRLGSKSVLSTCRAPDQPGCVL  
LHYSYDNGTIVKLLLEHYSYLNTHSPRIISVELPDDAKQFGIQRWMPQTHSSQGEDVW  
ADEILMTSVLNSISLDEFNLVEVTSIGFYLGNIQPYCGHDITLCFTCKSLASSM  
RYVETQSMQIGASVMTQFSLVGRGQKTYPHMDNQVKLEYSNHLGTLWHLVDECLPS  
MPCSQFTSASIIYHASEFTQWRVTVILPKMTQKATFRFWSQSYTAQDEWALDIIY  
IGQOCNPMCSHGSCDHGCRDQGVGQTECHPEAALPSTIMSDRENPSWSDWQEV  
IGGVVYPQGGVGVSSGSLYSFKAGKQLVSWDLTSMWDFVQYIIOIGESACN  
KPDREGVLLQYNNGGIQWHLLEAMFYSDRGKPRVYIELPAAAKTCTCTFRFWRW  
FSGEDYDQWAVDDIILISEKQKQIPVNPNTLPONFYEKPAFPDIPNGMSVWMLIAN  
EGMAKNDSTPISAMVFGKSGDRFAVTRDLTLKPGYVLOFKLNGISQSFSTAP  
VLIOYSHDAGSMWFLVKEGCFAPKAGCEGNSRELSEPTVYVYTGDFEWRVTVTAIP  
RLASAKTRFRIOESSOKNVPFGLDGVYISEPCPSYCSGHGDCISQCGDLGYT  
AAQTCVSNIPNHESEMFDRGKLSPLWKISGGVGTGCGTSLDGRSLYFNLGRRE  
ARTVPLDTRNIRLVQYIOGKTSITCICKPRARNEGVLVQYSDNGILWHLRLREL  
FLSFLPQIISIDLPREATPATAFRWQPHGKHSQAQWALDDVLVIGDSSQSTQFQD  
KFDMSIDLQANNYRIGQGVQDIDCLMSDALTALFTENIGKPRYAETWDFHVSASSYLFQ  
DKMSGCKSPFSATHVOLQYSLNNGKDMHPVTECPPTIGCVHTYESYTSERFQ  
WRVTVYLPATNSPRTRPRQANYTGADAWAIDNVLLASGCPWLCSGRGICDSGR  
CVCDRGFGGPGFVVPVPLPSILKDDFNGNLHPDLWPEVYGAERNGLNGETIKSGTSLI  
FKGEGRLMILSRDLCTNMYVQFSLRTAKGTPERSHSILLOSSINGVYWRLMDEF  
YFPQTSILFINVPLYSQAQTNATRLRWQPNNGKKEIWIIDDFILOGDNLNNPVM  
LTLDFDGRFNWFFYPGNIGLYCYPYSSKGAPEEDSAMVFSNIEGHSITTRDLS  
VENTIIIOFEINVGCTSDSSADPVRLEFSDFGATWHLHLLPLCYHSSLVSLSCST  
HHPSTIYAGTQGWREVHFHFKLHLCGSFRWYQGFYPAGSQPVTWAIQNVYIGP  
QCEEMCCGSGVNGTKICDPYSGPTCKISTKNPDLKDDFEGOLESDRFLMSGG  
KPSKCGIILSGNNLFNEDGLRMLVTRDLDLSHAREVQFMRGLCGKGVDPDRSOPV  
LQYSLNGLSLSLOEFLPSNNGVRYTIALEMPUKARSGDPLRWMPQSPENGHYS  
PWYDIIILGNSGVNLEDDFTLSDSRKWLHPGGTKMPVCGSTGDALVFIEKAST  
RVVYTDIAVNEDESLQIDFAASCYSTDCYALEYSDVLGLSHPLVRDCLPTNVE  
CSRYHLQRLVSLDTNFKWTRITPLPAYTRSOATRPRWHPAPDFKQOQWAIQNVYIG  
CDCLMCSGHRICIQGSCVCDQWQGLCYDEPETSILPTOLKDNFNAPSNQWLTNG  
GKLTVCGAVASGLAHFSGGCSRLLYVDLNLNAEFIOFYFMGCLITPNSRNOGV  
LLEYSVNGGITWLLMEIFYDQSKPGFVNILLPPDAKEITGRFRWQPRHDLQOND  
WAIQNVLIISGADQRTVMLDTFSSAPVQHPERSPADAGVRIAFDMEDRITSVEN  
VYFHDDCTYERFCDSPGVLACSHDREVYAVTHDLTPTENNFMQKTSVKGKVEK  
IAQNOIHVQSTDFGWSYVLPQCLPADPKCGTVSOPSVFPPTKGRITVPLPES  
LMGNVRRFREFYOKYSDVOWAIDNFXLPGCLDNGCHGDKLKEQCLICDPYSGPHYL  
THTLTKFLERDSEIKFDLWMSLEGGSTCEGLIAENTALIFGGSVROAITQDL  
DURKAFQWGRIGSNENNTSCHRVPCKRKGVLDDYSKDGGITWLLHEMDPFQKYS  
VRHDYILLPEGALTNTLRWQOPFVINSGLVYSGVERAQWALDNLILGGAENFSQL  
VDTFDDGSSHEENFSFYNAVRTAFGFCNPSFHLYPNKKDKTHALNSRELIIQP  
GYMOPFIYVGCATSCGDLHSMLEYTKDARSDWOLVQTOCLPSSNSIGCSPPQF  
HEATLYNANSSMRKITIQLPDHYVSSATQFRWIKGKETEKQSWAIDHYIGEACP  
RLCSGHGYCTTCAVICDESFGDDCVSFHELPSIKNFESARVETANWETIQGGA  
IGSGCQLAPYAHGSLXPNGCOIROAATKPLDLTRASKIMFVLQISTARMTQSCNSD  
LSGPHVTKRAVLQYVNNNGITHVIAHQPKDFEQAOQVRSYVNPLEARMKGYLLRW  
QPRHNGTGHQWALDHVEVLVSTRKQNTMMNFSRQHGLRHFYNNRRRSLRRYP"

BASE COUNT 2645 a 3029 c 2930 g 2583 t

ORIGIN

Query Match 9.1%; Score 32; DB 10; Length 11187;  
Best Local Similarity 100.0%; Pred. No. 6.1e-07;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 285 CGTGCAGTCCAGTGGAAACAGACAGCCTGC 316  
|||||

Db 1315 CGTGCAGTCCAGTGGAAACAGACAGCCTGC 1346  
|||||

RESULT 9  
AX410790

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1  
Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.  
Gene expression profiles in liver cancer

JOURNAL Patent: WO 0229103-A 3437 11-APR-2002;  
GENE LOGIC INC (US)  
FEATURES  
source  
Location/Qualifiers  
1. .11580  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/notes="EMBL/GenBank Accession No. U79716"  
BASE COUNT 3014 a 2696 c 2753 g 3116 t 1 others  
ORIGIN

Query Match 7.4%; Score 26; DB 6; Length 11580;  
Best Local Similarity 100.0%; Pred. No. 0.0019;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GAGAAATTAGCCCTTCCAATGT 233  
|||||

Db 1067 GAGAAATTAGCCCTTCCAATGT 1092  
|||||

RESULT 10  
HSU79716

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 11580)  
DeSilva, U., D'Arcangelo, G., Braden, V.V., Chen, J., Miao, G.G.,  
Curran, T. and Green, E.D.  
The human reelin gene: isolation, sequencing, and mapping on  
chromosome 7  
Genome Res. 7 (2), 157-164 (1997)  
97202106  
9049633  
2 (bases 1 to 11580)  
DeSilva, U., D'Arcangelo, G., Braden, V.V., Chen, J., Miao, G.G.,  
Curran, T. and Green, E.D.  
Direct Submission  
Submitted (26-NOV-1996) National Center for Human Genome Research,  
National Institutes of Health, 49 Convent Drive, MSC4431, Bethesda,  
MD 20892, USA  
FEATURES  
source  
Location/Qualifiers  
1. .11580  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="7"  
1. .11580  
/gene="RELN"  
176..10558  
/gene="RELN"  
/codon\_start=1  
/product="reelin"  
/db\_xref="GI:1743885"  
/translation="MERSGARQTFLLALLGATLRARAAAGYYPFRSPFFFLCTHHG  
ELEGGEGEVLISLHIAQNTYVYVPGQVHVITSTFTFDGLLVTLGLVTSVQASQ  
SIGGSAFGFGIMSDHQFQGNQPMCSVASHVSHLPTNLSFIWIAPPACTGCVNFMAT  
ATHRGQVIFDALAQQLCEGAPTVTVHPHLAEIHSDSIILRRDFTDSFVQLQNLPI  
WYFCNNECTGEGQAIMHGNVATFCPEYPRELITTLNTTASVQLQFSGISGRFS  
YSDPSIIYLYAKNSADWIKLRAPSNVSTIIHLYPEDAKGENVQFQWQENLR  
VGEVYACWALDNILINSARHQRVLESLDPDVTGNLWLPFGATVKHSCQDGNRIY  
FHNGSESENFATRDVLDSTEDIQOWSEERESEOPTGHDVILGAVTGTCSTIESGLS  
MYFLKDGKRLCTPSMDTGYGNLRFYVMGICDPGNSHENDIILYALIEGRKEHIT  
LDTLSSYSKYPSLVSVINPELQTPATKFCQRKNHQHNNRNWAVDFHVLPLPS  
TWSHMTQFINLGCCTHQPNSVSLFTSNHGRSLLHTECLPETCAGCPHFKCSGRG  
SSNTSGWNRITPLPNAALTRNTRIRWQTGPILGNMWAIDNVYIGPSCILKFCSGRG  
QCTRHGCKDCPDGSPACEMASOTPFMFISESGSRLSSYHNFTSIRCAEYVFCGV  
LASGKALVFNKDGRODLITSLFDSSQSRRFLOFTLRLGSKSVLSTCRAPDQPGCVLH  
YSYDNGITWKLLEHYSYLNTHSPRIISVELPDDAKQFGIQRWMPQTHSSQGEDVW

DEIIMTVLNFNSISLDFTNLVEVTSGLFYLGNVQPCYCGHDWTLCTGTGDSKLASSMRY  
VETQSGIYASNYOISFLVMGCGQKTYPHMDNQVLEYSTNHGLTWHLWQEBCLSPMP  
SCEQFTSAGIYHASTOETOMRRLVILLPOKTSWSSATFRWSSQYTYTAODWADLSYIG  
OCPNMGSGHSCDHCICRCDGOGYOGTECHPEAALPSTLMSDFENQNGWESDQWQVIG  
GEIVPEQCGISYSGSSLYFSKAKRQLVSWDLTSDTSDVDFVQFTIQIGGESACNKP  
DSREGEVLQIYNGNGIOHLLAEYMFSDFSKPRFVYLELPAATKPTCTFRFRWQPVF  
SEEDYDQWAVDIIILSEKOKIIPINPTLPONFYEKPAFDYPMQMSVLMLANEG  
MYKNETFCASAMIFGKSGDGRFAVTRDLTKPKYVLQKFLNIGCANQFSTAPVL  
LOYSHDAGSWFLVXGCGYPASAGKCEGNSRELSEPTMYHTGDFEEMTRITVIPS  
LASSKTRRWIOESSKKNVPPFGLDGVYISPCPSYCSGHGDCISGVCFCDLGYTAA  
OCTCVSNVHNEMEDPREGKLSPLWYKITGAQVGTGCGTLNDGKSLFNGPGKREAR  
TYPLDTRNLRLVQFTIQIGKSTGSLICINPRNEGLVVOISNDNGIMLHLLEDDEFM  
SLEPQIISIDLPQAKTAPATFRMWPQOHKSAQALDDVLLIGNMSSQTFQDDKE  
DGSIDLQANNYIQGGVDIDCLMDTLIFTENICKPRTAETDFHVSASTFLOFEM  
SMGCKSPFNHSVQLOSLNKGKDWHLTECVPTIGLHYTSSYIYTSERFQNW  
RTVYLPSTISIPRTFRWIOQANTYGVADSWAIDNVLASGPMWCSGKIGICDAGRCV  
CDRGEGGVCVPVPLPSLLKDDFNGLHPLDLPVYCGNERNLNGETIKSCTSLIFK  
GGGLRMLSRDLCNTIMVQSLRPIAKSTPERSHILLQFSISGGITWHLDEDFY  
POTTNIFLNVPLPAGTATNTRFLMQPYNNKKEEIVDDFTIDGNVNNVPMLL  
DTDFGPREDNFFYPGNGIGLYCPSYSSKAPEEDSAMVFSNEVGEHSITTRDLNV  
ENTIIQFEINQVSDSSADPVRELFSDRGATWHLPLCYHSSSHVSSLCSTEHH  
PSTYVAGTMOQRRREVHFGKHLGCSVRERWYOGFYFAGSQPVTAIDNVYIGQC  
EMCNGOGSCINGTKICDPYSGPTCKISTNPDLKDDFEGOLESDRELLMSGKP  
SKKGLSSGNLNFNEDGLRMLTDLDLSHAREVQFMRGLGCGVDPKRSQPVLL  
QYSLNGLSLQLEFLFSNSNVGYIALEIPLKARSSTRLRWQSPSENGHYFSPW  
VIDQILIGNISGNTVLEDDFTLDSRWKLLHGGTKMPVCGSTGDALVFIKASTRY  
VYSTDVAVNEDSFLOIDFAASCSTVDSYVAIELEYSDGLGSLWHLVRDCLPTNVECS  
RYHLORILVDFTNKTRITLPLPYTRSOATFRWHPAPDFKQOTWAIIDNVYIGDG  
CIDMGSGHRCIOGNCVCDDEOAGGLYCDPETSLOTQKDNENRAPSOWITVNGGK  
LSTVGVASGHALHFGSGCRLVTVLDNLNAEFTQYFMIGLITPNNRNQGVLL  
EYSVNGGITWLLMEIFYDOYSKPGFVNILLPPDAKETATFRFWMOPRHGDQNDWA  
IDNVLIGSADQRTVMDLTFSSAPVQHPERSPADAGPVGRIAFDMEDKTSVNEHWL  
FHDDCTVERFCDSPGMLCGSHDREYAVTHDLTPTEGIMQFKRISYCKVSEKIA  
QNOIHVQSTDFGVSNNYLVPOCLPADPKCSVSQPSVFFPTKWKRYITPLPESLV  
GNPFRFYQYSDMOWADNFYLGPCDNCRGHDCLEOCICDPGSGPNCYLTH  
TLKTELKERDESEEIKPOLMNSLEGSTCEGILAEADTALFYGSGTVQAOYQDLDL  
RKAFLQWGRIGSENMTSCHRPICRKEGVLLDSTDGKITWLLHENDYOKYISVR  
HDYILDPEDALINTRLRWQPFVINSIVGVSVERAQAQWALDNLIGGAETINPSQLVD  
TFDDETSHEENSYFPNAVRTAGFCGNPSFHYWPKKKDKTHNALSRELIIQPGY  
MQQFKIVGCEATSCGDLHSMLEYTKDARSWSQVLOTQCLPSSNSIGCSPFQFHE  
ATLYNSVSSSKWKRTIOLPDHVSSTATOFRTWIKQETEKOSWAIIDHVIAGEAPKL  
CSHGVCYTTGATCIDCESFGQDCCSVESHDLPYIKDNFESARVETANWETIOGGVIG  
SCGGQLAPYAHGDSLYFNGCQIQROATRPDLITRASKINFYVLIQISMSOTDSCNSDLS  
GPHAVDKVLLQIYSVNGNITHVIAHQHPQDFTQAOVSIVNPLEARMKGVLLRWQP  
RINGTCHDQWALDHVEVLVSTRKQNMNFRSQRLHFYNNRRSLRRYP"

BASE COUNT 3014 a 2696 c 2753 g 3116 t 1 others

Query Match 7.4%; Score 26; DB 9; Length 11580;  
Best Local Similarity 100.0%; Pred. No. 0.0019;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 208 GAGAAATAGAGCCCTTCCCAATGT 233  
|||||

Db 1067 GAGAAATAGAGCCCTTCCCAATGT 1092

RESULT 11

G30936 11580 bp DNA linear STS 28-SEP-1998

LOCUS SW52926 Eric D. Green Homo sapiens STS genomic, sequence tagged

DEFINITION site.

ACCESSION G30936

VERSION G30936.1 GI:1923209

KEYWORDS STS.

SOURCE Homo sapiens.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 11580)  
Bouffard,G.G., Iyer,L.M., Idol,J.R., Braden,V.V., Cunningham,A.F.,  
Weintraub,L.A., Mohr-Tidwell,R.M., Peluso,D.C., Fulton,R.S.,  
Leckie,M.P. and Green,E.D.

TITLE A collection of 1814 human chromosome 7-specific STSs  
JOURNAL Genome Res. 7 (1), 59-64 (1997)  
MEDLINE 97189344  
PUBMED 9037602  
REFERENCE 2 (bases 1 to 11580)  
AUTHORS Green,E.D.  
TITLE Human chromosome 7 STSs (1997)  
JOURNAL Unpublished (1997)  
COMMENT On Apr 3, 1997 this sequence version replaced gi:1706935.  
Synonyms: RELN  
GDB\_DSEG: RELN

Contact: Eric D. Green  
Genome Technology Branch  
National Human Genome Research Institute/NIH  
49 Convent Dr., MSC4431, Bldg. 49, Rm. 2A08, Bethesda, MD 20892  
Tel: 3014020201  
Fax: 3014024735  
Email: egreen@nhgri.nih.gov  
Primer A: CTCTACCTCTCTGAGGACGCCAA  
Primer B: AAGCAGTTGCTGTGTCCACTG  
STS size: 183  
PCR Profile:

Prisoak: 0 degrees C for 0.00 minute(s)  
Denaturation: 92 degrees C for 0.17 minute(s)  
Annealing: 55 degrees C for 1.00 minute(s)  
Polymerization: 72 degrees C for 1.00 minute(s)  
PCR Cycles: 35  
Thermal Cycler: PerkinElmer 9600  
Protocol:  
Template: 30-100 ng  
Primer: each 1 uM  
dNTPs: each 200 uM  
Tag Polymerase: 0.05 units/uL  
Total Vol: 10 uL

Buffer:  
MgCl2: 1.5 mM  
KCl: 100 mM  
Tris-HCl: 10 mM  
NH4Cl: 5 mM  
pH: 8.6

This STS was developed from sequence determined by another investigator. See GenBank record: U79716 For additional information about the NHGRI chromosome 7 mapping project, see <http://www.nhgri.nih.gov/DIR/STB/CHR7>. Also see Genomics 11:548-64 (1991) [MUID=92128937].

FEATURES  
source location/Qualifiers  
1..11580  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="7"  
/clone\_lib="Eric D. Green"  
1..11580  
/gene="RELN"  
1112..1294  
/gene="RELN"  
1112..1134  
/gene="RELN"  
primer\_bind complement(1272..1294)  
primer\_bind 3014 a 2696 c 2753 g 3116 t 1 others  
BASE COUNT 3014 a 2696 c 2753 g 3116 t 1 others  
ORIGIN

Query Match 7.4%; Score 26; DB 11; Length 11580;  
Best Local Similarity 100.0%; Pred. No. 0.0019;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 208 GAGAAATAGAGCCCTTCCCAATGT 233  
|||||

Db 1067 GAGAAATAGAGCCCTTCCCAATGT 1092

RESULT 12

```

G30938      11580 bp   DNA       linear   STS 28-SEP-1998
LOCUS      SWS3176 Eric D. Green Homo sapiens STS genomic, sequence tagged
DEFINITION site.
ACCESSION  G30938
VERSION    G30938.1 GI:1923211
KEYWORDS   STS.
SOURCE     Homo sapiens.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 11580)
AUTHORS   Bouffard,G.G., Iyer,L.M., Idol,J.R., Braden,V.V., Cunningham,A.F.,
            Weintraub,L.A., Mohr-Tidwell,R.M., Peluso,D.C., Fulton,R.S.,
            Leckie,M.P. and Green,E.D.
TITLE     A collection of 1814 human chromosome 7-specific STSs
JOURNAL   Genome Res. 7 (1), 59-64 (1997)
MEDLINE   97189344
PUBMED    9037602
REFERENCE  2 (bases 1 to 11580)
AUTHORS   Green,E.D.
TITLE     Human chromosome 7 STSs (1997)
JOURNAL   Unpublished (1997)
COMMENT   On Apr 3, 1997 this sequence version replaced gi:1706937.
SYNONYMS  RELN
GDB_DSEC  RELN
Contact:  Eric D. Green
Genome Technology Branch
National Human Genome Research Institute/NIH
49 Convent Dr., MSC4431, Bldg. 49, Rm. 2A08, Bethesda, MD 20892
Tel: 3014020201
Fax: 3014024735
Email: egreen@nhgri.nih.gov
Primer A: TGTGGCTGATTACTAACAG
Primer B: TGGGCTTTATAGCTTG
STS size: 238
PCR Profile:
    Presoak: 0 degrees C for 0.00 minute(s)
    Denaturation: 92 degrees C for 0.17 minute(s)
    Annealing: 50 degrees C for 1.00 minute(s)
    Polymerization: 72 degrees C for 1.00 minute(s)
    PCR Cycles: 35
    Thermal Cycler: PerkinElmer 9600
Protocol:
    Template: 30-100 ng
    Primer: each 1 uM
    dNTPs: each 200 uM
    Taq Polymerase: 0.05 units/ul
    Total Vol: 10 ul

Buffer:    MgCl2: 1.5 mM
           KCl: 100 mM
           Tris-HCl: 10 mM
           NH4Cl: 5 mM
           pH: 8.6

This STS was developed from sequence determined by another
investigator. See GenBank record: U79716 For additional
information about the NHGRI chromosome 7 mapping project, see
http://www.nhgri.nih.gov/DIR/STS/CHR7. Also see Genomics
11:548-64 (1991) [MUID=92128937].
FEATURES   Location/Qualifiers
            source
            1..11580
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /map="7"
                /clone_lib="Eric D. Green"
            gene
            1..11580
                /gene="RELN"
            STS
            11314..11551
                /gene="RELN"
primer_bind 11314..11333

```

---

```

/ gene="RELN"
primer_bind 3014 a 2696 c 2753 g 3116 t 1 others
BASE COUNT 3014 a 2696 c 2753 g 3116 t 1 others
ORIGIN
Query Match 7.4%; Score 26; DB 11; Length 11580;
Best Local Similarity 100.0%; Pred No. 0.0019;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 208 GAGAAATTAGAGCCCTTCCAATGT 233
Db 1067 GAGAAATTAGAGCCCTTCCAATGT 1092
RESULT 13
AC119906/c
LOCUS      Mus musculus clone RP24-239D13, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC119906
ACCESSION  AC119906
VERSION    AC119906.1 GI:20389564
KEYWORDS   HTG; HTGS_PHASE0.
SOURCE     Mus musculus.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 70905)
AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE     Mus musculus, clone RP24-239D13
JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 70905)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Illiev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Lacombe,K., Lamazares,R.,
            Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Maneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Triggilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
            Direct Submission
            Submitted (02-MAY-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            All repeats were identified using RepeatMasker:
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            Smith, A.F.A. & Green, P. (1996-1997)
            ----- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ----- Project Information
            Center project name: L25284
            Center clone name: 239_D_13
            -----
            * NOTE: This record contains 89 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows

```

\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1  
692 791: gap of 100 bp  
792 1481: contig of 690 bp in length  
1482 1581: gap of 100 bp  
1582 2255: contig of 714 bp in length  
2296 2395: gap of 100 bp  
2396 3014: contig of 619 bp in length  
3015 3114: gap of 100 bp  
3115 3850: contig of 736 bp in length  
3851 3950: gap of 100 bp  
3951 4619: contig of 669 bp in length  
4620 4719: gap of 100 bp  
4720 5407: contig of 688 bp in length  
5408 5507: gap of 100 bp  
5508 6196: contig of 689 bp in length  
6197 6296: gap of 100 bp  
6297 7009: contig of 713 bp in length  
7010 7109: gap of 100 bp  
7110 7805: contig of 696 bp in length  
7806 7905: gap of 100 bp  
7906 8615: contig of 710 bp in length  
8616 8715: gap of 100 bp  
8716 9429: contig of 714 bp in length  
9430 9529: gap of 100 bp  
9530 10249: contig of 720 bp in length  
10250 10349: gap of 100 bp  
10350 11059: contig of 710 bp in length  
11060 11159: gap of 100 bp  
11160 11856: contig of 697 bp in length  
11857 11956: gap of 100 bp  
11957 12632: contig of 676 bp in length  
12633 12732: gap of 100 bp  
12733 13425: contig of 693 bp in length  
13426 13525: gap of 100 bp  
13526 14193: contig of 668 bp in length  
14194 14293: gap of 100 bp  
14294 14959: contig of 666 bp in length  
14960 15059: gap of 100 bp  
15060 15752: contig of 693 bp in length  
15753 15852: gap of 100 bp  
15853 16545: contig of 693 bp in length  
16546 16645: gap of 100 bp  
16646 17336: contig of 691 bp in length  
17337 17436: gap of 100 bp  
17437 18129: contig of 693 bp in length  
18130 18229: gap of 100 bp  
18230 18944: contig of 715 bp in length  
18945 19044: gap of 100 bp  
19045 19763: contig of 719 bp in length  
19764 19863: gap of 100 bp  
19864 20586: contig of 723 bp in length  
20587 20686: gap of 100 bp  
20687 21421: contig of 735 bp in length  
21422 21521: gap of 100 bp  
21522 22211: contig of 690 bp in length  
22212 22311: gap of 100 bp  
22312 22962: contig of 651 bp in length  
22963 23062: gap of 100 bp  
23063 23748: contig of 686 bp in length  
23749 23848: gap of 100 bp  
23849 24535: contig of 687 bp in length  
24536 24635: gap of 100 bp  
24636 25329: contig of 694 bp in length  
25330 25429: gap of 100 bp  
25430 26129: contig of 700 bp in length  
26130 26229: gap of 100 bp  
26230 26929: contig of 700 bp in length  
26930 27029: gap of 100 bp

27030 27732: contig of 703 bp in length  
27733 27832: gap of 100 bp  
27833 28548: contig of 716 bp in length  
28549 28648: gap of 100 bp  
28649 29367: contig of 719 bp in length  
29368 29467: gap of 100 bp  
29468 30168: contig of 701 bp in length  
30169 30268: gap of 100 bp  
30269 30961: contig of 693 bp in length  
30962 31061: gap of 100 bp  
31062 31759: contig of 698 bp in length  
31760 31859: gap of 100 bp  
31860 32555: contig of 696 bp in length  
32556 32655: gap of 100 bp  
32656 33321: contig of 666 bp in length  
33322 33421: gap of 100 bp  
33422 34116: contig of 695 bp in length  
34117 34216: gap of 100 bp  
34217 34907: contig of 691 bp in length  
34908 35007: gap of 100 bp  
35008 35706: contig of 699 bp in length  
35707 35806: gap of 100 bp  
35807 36516: contig of 710 bp in length  
36517 36616: gap of 100 bp  
36617 37312: contig of 696 bp in length  
37313 37412: gap of 100 bp  
37413 38141: contig of 729 bp in length  
38142 38241: gap of 100 bp  
38242 38967: contig of 726 bp in length  
38968 39067: gap of 100 bp  
39068 39760: contig of 693 bp in length  
39761 39860: gap of 100 bp  
39861 40551: contig of 691 bp in length  
40552 40651: gap of 100 bp  
40652 41332: contig of 681 bp in length  
41333 41432: gap of 100 bp  
41433 42115: contig of 683 bp in length  
42116 42215: gap of 100 bp  
42216 42906: contig of 691 bp in length  
42907 43006: gap of 100 bp  
43007 43718: contig of 712 bp in length  
43719 43818: gap of 100 bp  
43819 44518: contig of 700 bp in length  
44519 44618: gap of 100 bp  
44619 45321: contig of 703 bp in length  
45322 45421: gap of 100 bp  
45422 46126: contig of 705 bp in length  
46127 46226: gap of 100 bp  
46227 46946: contig of 720 bp in length  
46947 47046: gap of 100 bp  
47047 47738: contig of 692 bp in length  
47739 47838: gap of 100 bp  
47839 48523: contig of 685 bp in length  
48524 48623: gap of 100 bp  
48624 49319: contig of 696 bp in length  
49320 49419: gap of 100 bp  
49420 50106: contig of 687 bp in length  
50107 50206: gap of 100 bp  
50207 50888: contig of 682 bp in length  
50889 50988: gap of 100 bp  
50989 51681: contig of 693 bp in length  
51682 51781: gap of 100 bp  
51782 52476: contig of 695 bp in length  
52477 52576: gap of 100 bp  
52577 53300: contig of 724 bp in length  
53301 53400: gap of 100 bp  
53401 54093: contig of 693 bp in length  
54094 54193: gap of 100 bp  
54194 54893: contig of 700 bp in length  
54894 54993: gap of 100 bp

Query Match  
Best Local Similarity

6.8%; Score 24; DB 2; Length 70905;  
100.0%; Pred. No. 0.027;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 99 ATCTGTCTCCAGTTTCCTCATTTG 122  
|||||CCCCCCCCCCCCCCCCCCCC|

Db 60522 ATCTGTCTCCAGTTTCCTCATTTG 60499

RESULT 14  
AL590143  
LOCUS  
DEFINITION  
ACCESSION  
AL590143  
VERSION  
AL590143.4  
KEYWORDS  
GI:14031120  
SOURCE  
HTG.  
ORGANISM  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 73070)  
Dunn.M.  
Direct Submission  
Submitted (10-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk  
On May 14, 2001 this sequence version replaced gi:13751595.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em.; EMBL; Sw.;  
SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information on the WORMPEP  
database can be found at  
http://www.sanger.ac.uk/Projects/C-elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping  
Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr6  
RP11-132M7 is from the library RPCI-11.1 constructed by the group  
of Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pBACe3.6  
IMPORTANT: This sequence is not the entire insert of clone  
RP11-132M7. It may be shorter because we sequence overlapping  
sections only once, except for a 100 base overlap.  
The true left end of clone RP1-33L1 is at 72971 in this sequence.  
The true right end of clone RP11-533L1 is at 100 in this sequence.

FEATURES  
source  
Location/Qualifiers  
1..73070  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/clone="RP11-132M7"  
/clone\_lib="RPCI-11.1"  
41..418  
/note="MT1A1 repeat: matches 8. .365 of consensus"  
853..1165  
/note="MT1A1 repeat: matches 1. .321 of consensus"  
4023..4373  
/note="MT1A1 repeat: matches 1. .362 of consensus"  
5311..5463  
/note="MER58A repeat: matches 72. .224 of consensus"  
5533..5743

repeat\_region  
/note="MER5A repeat: matches 2. .188 of consensus"  
5755..5949  
/note="MT1H repeat: matches 127. .323 of consensus"  
7581..7766  
/note="MT1A2 repeat: matches 1. .185 of consensus"  
7767..7800  
/note="17 copies 2 mer ct 94% conserved"  
7820..7995  
/note="MT1A2 repeat: matches 173. .361 of consensus"  
9102..9435  
/note="AluX repeat: matches 1. .311 of consensus"  
10262..10404  
/note="MIR repeat: matches 97. .262 of consensus"  
11515..11681  
/note="L1PAL5 repeat: matches 5990. .6157 of consensus"  
11974..12101  
/note="MIR repeat: matches 117. .257 of consensus"  
12390..12696  
/note="AluSg repeat: matches 1. .306 of consensus"  
13603..13899  
/note="L1PAL3 repeat: matches 5860. .6156 of consensus"  
15223..15252  
/note="15 copies 2 mer ac 100% conserved"  
16988..17111  
/note="62 copies 2 mer ca 71% conserved"  
17113..17208  
/note="48 copies 2 mer ca 70% conserved"  
17462..17516  
/note="L2 repeat: matches 2647. .2702 of consensus"  
17637..17736  
/note="50 copies 2 mer aa 61% conserved"  
18105..18277  
/note="MER41B repeat: matches 258. .430 of consensus"  
18772..19623  
/note="L1PAL3 repeat: matches 5328. .6156 of consensus"  
19628..20044  
/note="HAL1 repeat: matches 198. .632 of consensus"  
2047..20108  
/note="31 copies 2 mer aa 71% conserved"  
20345..20549  
/note="MIR repeat: matches 2. .236 of consensus"  
20585..20791  
/note="HAL1 repeat: matches 1286. .1497 of consensus"  
20795..20841  
/note="L2 repeat: matches 2662. .2709 of consensus"  
21022..21340  
/note="AluJb repeat: matches 1. .311 of consensus"  
23079..23396  
/note="AluYb8 repeat: matches 1. .318 of consensus"  
23575..23778  
/note="L2 repeat: matches 1254. .1475 of consensus"  
23788..23975  
/note="L2 repeat: matches 2541. .2700 of consensus"  
24047..24343  
/note="AluJb repeat: matches 5. .307 of consensus"  
24568..24870  
/note="AluSc repeat: matches 1. .309 of consensus"  
25992..26048  
/note="L2 repeat: matches 2678. .2734 of consensus"  
26547..26726  
/note="L2 repeat: matches 1999. .2181 of consensus"  
27685..27758  
/note="L2 repeat: matches 1777. .1850 of consensus"  
28576..28791  
/note="MIR repeat: matches 21. .262 of consensus"  
28798..29000  
/note="MIR repeat: matches 57. .262 of consensus"  
29611..29785  
/note="MIR repeat: matches 20. .234 of consensus"  
29894..30003  
/note="Tigger3(Golem) repeat: matches 2929. .3038 of  
consensus"  
30130..30209

repeat\_region /note="Tigger3(Golem) repeat: matches 1. .80 of consensus"  
repeat\_region 30330. .30652  
repeat\_region /note="LIME2 repeat: matches 5817. .6287 of consensus"  
repeat\_region 30812. .31142  
repeat\_region /note="AlusX repeat: matches 1. .303 of consensus"  
repeat\_region 32108. .32297  
repeat\_region /note="L1 repeat: matches 2822. .3028 of consensus"  
repeat\_region 32300. .33017  
repeat\_region /note="MER50 repeat: matches 1. .711 of consensus"  
repeat\_region 33115. .33682  
repeat\_region /note="L1 repeat: matches 3078. .3649 of consensus"  
repeat\_region 33789. .34325  
repeat\_region /note="MLT2B repeat: matches 2. .448 of consensus"  
repeat\_region 34391. .34544  
repeat\_region /note="L1 repeat: matches 3828. .3985 of consensus"  
repeat\_region 34545. .34856  
repeat\_region /note="AluY repeat: matches 1. .311 of consensus"  
repeat\_region 34857. .35048  
repeat\_region /note="L1 repeat: matches 3985. .4164 of consensus"  
repeat\_region 35415. .35601  
repeat\_region /note="MIR repeat: matches 21. .200 of consensus"  
repeat\_region 35701. .35836  
repeat\_region /note="L2 repeat: matches 2585. .2735 of consensus"  
repeat\_region 37327. .38045  
repeat\_region /note="L2 repeat: matches 2585. .2745 of consensus"  
repeat\_region 39417. .39539  
repeat\_region /note="MIR repeat: matches 19. .147 of consensus"  
repeat\_region 40146. .40185  
repeat\_region /note="20 copies 2 mer gt 85% conserved"  
repeat\_region 41086. .41217  
repeat\_region /note="L2 repeat: matches 1421. .1557 of consensus"  
repeat\_region 41236. .41369  
repeat\_region /note="MER63 repeat: matches 1. .766 of consensus"  
repeat\_region 41831. .42070  
repeat\_region /note="L1M4C repeat: matches 1617. .1860 of consensus"  
repeat\_region 42121. .42155  
repeat\_region /note="L1P4A repeat: matches 6112. .6146 of consensus"  
repeat\_region 42156. .42656  
repeat\_region /note="L1M4C repeat: matches 1072. .1565 of consensus"  
repeat\_region 43386. .43486  
repeat\_region /note="L1M49 repeat: matches 6210. .6308 of consensus"  
repeat\_region 45345. .45536  
repeat\_region /note="MIR repeat: matches 55. .246 of consensus"  
repeat\_region 46312. .46359  
repeat\_region /note="24 copies 2 mer ta 81% conserved"  
repeat\_region 48178. .48223  
repeat\_region /note="23 copies 2 mer ac 100% conserved".  
repeat\_region 48644. .48746  
repeat\_region /note="L2 repeat: matches 2159. .2262 of consensus"  
repeat\_region 49582. .49887  
repeat\_region /note="AlusQ repeat: matches 1. .299 of consensus"  
repeat\_region 49888. .56031  
repeat\_region /note="L1P4M repeat: matches 4. .6144 of consensus"  
repeat\_region 56756. .56848  
repeat\_region /note="L2 repeat: matches 2598. .2700 of consensus"  
repeat\_region 58357. .58412  
repeat\_region /note="28 copies 2 mer tg 87% conserved"  
repeat\_region 58586. .58669  
repeat\_region /note="MIR repeat: matches 4. .92 of consensus"  
repeat\_region 58750. .58999  
repeat\_region /note="MIR repeat: matches 6. .252 of consensus"  
repeat\_region 60707. .60846  
repeat\_region /note="MER3 repeat: matches 53. .169 of consensus"

Query Match : 6.0%; Score 21; DB 9; Length 73070;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGCAGTGTGGCACCATCATG 21  
|||||  
Db 2245 GAGCAGTGTGGCACCATCATG 2265

RESULT 15  
AC015229 98873 bp DNA linear HTG 16-NOV-1999  
LOCUS Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*, in ordered  
DEFINITION pieces.  
AC015229  
VERSION AC015229.1 GI:6436106  
KEYWORDS HTG; HTGS\_PHASE2.  
SOURCE Drosophila melanogaster.  
ORGANISM Drosophila melanogaster.  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 98873)  
AUTHORS Adams, M. and Venter, J.C.  
TITLE Direct Submission  
JOURNAL Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,  
Rockville, MD, USA  
COMMENT This sequence was identified as CDN:10214184 by the submitter.  
For further information on this sequence e-mail to fly@celera.com.  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
FEATURES  
source  
1. .98873  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
BASE COUNT 29559 a 19891 c 19792 g 29631 t  
ORIGIN  
Query Match 6.0%; Score 21; DB 2; Length 98873;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 188 CCGCTGATTGGATTCAGCTGG 208  
|||||  
Db 57715 CCGCTGATTGGATTCAGCTGG 57735  
Search completed: November 6, 2002, 17:40:29  
Job time : 3116 secs



SQ Sequence 7215 BP; 2335 A; 1369 C; 1559 G; 1952 T; 0 other;  
 Query Match 5.4%; Score 19; DB 22; Length 7215;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 286 GTGCAGTTCGACGTGAAC 304  
 |||||  
 DB 3906 GTGCAGTTCGACGTGAAC 3924

RESULT 14  
 ABL40174  
 ID ABL40174 standard; DNA; 18 BP.  
 XX  
 AC ABL40174;  
 XX  
 DT 21-MAY-2002 (first entry)  
 XX

Mouse reelin protein CR-50 epitope region PCR primer SEQ ID NO:11.

Mouse; reelin protein CR-50 epitope region; elucidation: neuron;  
 cerebral disturbance; reelin protein; neuroprotective; PCR primer: ss.

OS Mus musculus.  
 PN JP2002017361-A.  
 XX

22-JAN-2002.

04-JUL-2000; 2000JP-0202801.

04-JUL-2000; 2000JP-0202801.

(RIKA) RIKAGAKU KENKYUSHO.

WPI; 2002-221707/28.

Reelin protein CR-50 epitope region, useful for diagnosis and treatment  
 of cerebral disturbance

Example 2; Page 7; 16pp; Japanese.

The present invention describes the mouse reelin protein CR-50 epitope  
 region, which contains the CR-50 antibody recognition site and is free  
 from F-spondin domains and repetitive sites. Also described are: (1) an  
 expression vector comprising a polynucleotide encoding a reelin protein  
 epitope region; (2) host cells with transfected the expression vector;  
 (3) polypeptides prepared by culture of the host cells; and (4)  
 polynucleotides comprising the 351 base sequence given in ABL40165 which  
 encodes the 117 amino acid sequence given in AB06244; and (5) use of  
 the polynucleotide for diagnosis and/or treatment of diseases caused by  
 abnormal positioning of neural cells, and stimulation of association of  
 reelin protein. The mouse reelin protein CR-50 epitope region has  
 neuroprotective activity, and can be used in the diagnosis and treatment  
 of cerebral disturbance due to an abnormal reelin gene and positioning  
 of neurons. The present sequence represents a PCR primer for the mouse  
 reelin protein CR-50 epitope region, which is used in an example from  
 the present invention.

SQ Sequence 18 BP; 4 A; 5 C; 6 G; 3 T; 0 other;

Query Match 5.1%; Score 18; DB 24; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 49;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGCAGTGTGACACATC 18  
 |||||  
 DB 1 GAGCAGTGTGACACATC 18

RESULT 15  
 AAA70701

ID AAA70701 standard; cDNA; 975 BP.

XX AAA70701;

DT 11-DEC-2000 (first entry)

Human interleukin 6 receptor cDNA.

Human; interleukin-6 receptor; fungus: Pichia pastoris; PCR primer: ss;  
 expression vector; immunoglobulin-like region; cytokine receptor region.

Homo sapiens.

Key Location/Qualifiers  
 CDS 1..975

FT /tag= a

FT /partial

FT /product= "human IL-6R amino acids 20-344"

FT /note= "no start or stop codon is given at the 5' or

3' ends of the sequence"

FT /transl\_except= (pos:169..171,aa:Glu)

FT /transl\_except= (pos:841..843,aa:Gly)

PN JP2000157280-A.

13-JUN-2000.

26-NOV-1998; 98JP-0335464.

26-NOV-1998; 98JP-0335464.

(TOYJ) TOSOH CORP.

WPI; 2000-468203/41.

P-PSDB; AAB15389.

Yeasts transformed with IL-6 receptor gene

Example 1; Page 6-8; 10pp; Japanese.

The invention relates to the production of human interleukin-6 receptor  
 (IL-6R) protein in the fungus Pichia pastoris. The fungus was  
 transformed with an expression vector (pPIC9-A20HL) containing a gene  
 encoding an IL-6R protein having an immunoglobulin-like region and  
 cytokine receptor region and spanning amino acids from Leu20-Ala323. The  
 human IL-6R region was PCR amplified using the primers AAA70702-A70703.  
 This sequence represents the coding region for the IL-6R protein of the  
 invention.

SQ Sequence 975 BP; 208 A; 289 C; 290 G; 188 T; 0 other;

Query Match 5.1%; Score 18; DB 21; Length 975;  
 Best Local Similarity 100.0%; Pred. No. 45;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 CCAGCATCACTGTGTCAAT 172  
 |||||  
 DB 762 CCAGCATCACTGTGTCAAT 779

Search completed: November 6, 2002, 16:48:34  
 Job time: 314 secs

```
XX Key Location/Qualifiers
FH CDS 1..4593
FT /*tag= a
FT /partial
FT /product "Human protein, SEQ ID 6"
FT /note= "No start codon given"
XX
XX
XX WO200127270-A1.
XX
XX PD 19-APR-2001.
XX
XX PE 25-AUG-2000; 2000WO-JP05756.
XX
XX PR 08-OCT-1999; 99JP-0288738.
XX PR 08-OCT-1999; 99JP-0288739.
XX PR 19-APR-2000; 2000JP-0123721.
XX
XX PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX
XX PI Taga T, Kimura N;
XX
XX DR WPI; 2001-282027/29.
XX
XX DR P-PSDB; AAB84882.
XX
XX PT DNA Sequences encoding YS68 protein involved in early hematopoiesis -
XX
XX PS Disclosure; Page 90-105; 245pp; Japanese.
XX
XX CC The present invention relates to novel DNA sequences encoding YS68
XX CC protein, which is involved in early hematopoiesis. The protein, YS68, is
XX CC involved in early haematopoiesis and the gene may be useful as a factor
XX CC for controlling it, or as a tool in development of related drugs. The
XX CC protein has potential use against tumours. The present sequence was used
XX CC in the present invention.
XX
XX SQ Sequence 4883 BP; 1699 A; 938 C; 982 G; 1264 T; 0 other;

Query Match 5.4%; Score 19; DB 22; Length 4883;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 286 GTGCAGTTCAGTGGAAAC 304
DB 1574 GTGCAGTTCAGTGGAAAC 1592

RESULT 12
AAH19497
ID AAH19497 standard; DNA: 7215 BP.
XX
XX AC AAH19497;
XX
XX DT 03-AUG-2001 (first entry)
XX
XX DE Human coding sequence, SEQ ID 13.
XX
XX DE Human coding sequence, SEQ ID 13.
XX
XX KW Human; YS68 protein; early haematopoiesis; tumour; ss.
XX
XX OS Homo sapiens.
XX
XX FH Key Location/Qualifiers
FT CDS 125..6925
FT /*tag= a
FT /product "Human protein, SEQ ID 14"
XX
XX PN WO200127270-A1.
XX
XX PD 19-APR-2001.
XX
XX PE 25-AUG-2000; 2000WO-JP05756.
XX
XX PR 08-OCT-1999; 99JP-0288738.
```

```
PR 08-OCT-1999; 99JP-0288739.
PR 19-APR-2000; 2000JP-0123721.
XX
XX PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX
XX PI Taga T, Kimura N;
XX
XX DR WPI; 2001-282027/29.
XX
XX DR P-PSDB; AAB84885.
XX
XX PT DNA Sequences encoding YS68 protein involved in early hematopoiesis -
XX
XX PS Claim 1; Page 180-202; 245pp; Japanese.
XX
XX CC The present invention relates to novel DNA sequences encoding YS68
XX CC protein, which is involved in early hematopoiesis. The protein, YS68, is
XX CC involved in early haematopoiesis and the gene may be useful as a factor
XX CC for controlling it, or as a tool in development of related drugs. The
XX CC protein has potential use against tumours. The present sequence was used
XX CC in the present invention.
XX
XX SQ Sequence 7215 BP; 2335 A; 1369 C; 1559 G; 1952 T; 0 other;

Query Match 5.4%; Score 19; DB 22; Length 7215;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 286 GTGCAGTTCAGTGGAAAC 304
DB 3906 GTGCAGTTCAGTGGAAAC 3924

RESULT 13
AAH19498
ID AAH19498 standard; DNA: 7215 BP.
XX
XX AC AAH19498;
XX
XX DT 03-AUG-2001 (first entry)
XX
XX DE Human coding sequence, SEQ ID 15.
XX
XX DE Human coding sequence, SEQ ID 15.
XX
XX KW Human; YS68 protein; early haematopoiesis; tumour; ss.
XX
XX OS Homo sapiens.
XX
XX OS Homo sapiens.
XX
XX PN WO200127270-A1.
XX
XX PD 19-APR-2001.
XX
XX PE 25-AUG-2000; 2000WO-JP05756.
XX
XX PR 08-OCT-1999; 99JP-0288738.
XX PR 08-OCT-1999; 99JP-0288739.
XX PR 19-APR-2000; 2000JP-0123721.
XX
XX PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX
XX PI Taga T, Kimura N;
XX
XX DR WPI; 2001-282027/29.
XX
XX PT DNA Sequences encoding YS68 protein involved in early hematopoiesis -
XX
XX PS Disclosure; Page 220-242; 245pp; Japanese.
XX
XX CC The present invention relates to novel DNA sequences encoding YS68
XX CC protein, which is involved in early hematopoiesis. The protein, YS68, is
XX CC involved in early haematopoiesis and the gene may be useful as a factor
XX CC for controlling it, or as a tool in development of related drugs. The
XX CC protein has potential use against tumours. The present sequence was used
XX CC in the present invention.
XX
```

CC of neurons. The present sequence represents a PCR primer for the mouse  
CC reelin protein CR-50 epitope region, which is used in an example from  
CC the present invention.

XX  
SQ Sequence 22 BP; 5 A; 9 C; 6 G; 2 T; 0 other;

Query Match 6.3%; Score 22; DB 24; Length 22;

Best Local Similarity 100.0%; Pred. No. 0.47;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 330 GTATGAGCCTGCTGGCCCTG 351

Db 22 GTATGAGCCTGCTGGCCCTG 1

RESULT 9

ABLA0173

ID ABLA0173 standard; DNA; 21 BP.

XX  
XX ABLA0173;

XX 21-MAY-2002 (first entry)

DE Mouse reelin protein CR-50 epitope region PCR primer SEQ ID NO:10.

XX  
KW Mouse; reelin protein CR-50 epitope region; elucidation; neuron;

XX cerebral disturbance; reelin protein; neuroprotective; PCR primer; ss.

OS Mus musculus.

XX  
XX JP2002017361-A.

XX  
XX 22-JAN-2002.

XX  
XX 04-JUL-2000; 2000JP-0202801.

XX  
XX 04-JUL-2000; 2000JP-0202801.

XX  
XX (RIKA ) RIKAGAKU KENKYUSHO.

XX  
XX WPI; 2002-221707/28.

PT Reelin protein CR-50 epitope region, useful for diagnosis and treatment  
PT of cerebral disturbance -

XX  
PS Example 1; Page 6; 16pp; Japanese.

CC The present invention describes the mouse reelin protein CR-50 epitope  
CC region, which contains the CR-50 antibody recognition site and is free  
CC from F-spondin domains and repetitive sites. Also described are: (1) an  
CC expression vector comprising a polynucleotide encoding a reelin protein  
CC epitope region; (2) host cells with transfected the expression vector;  
CC (3) polypeptides prepared by culture of the host cells; and (4)  
CC polynucleotides comprising the 351 base sequence given in ABLA0165 which  
CC encodes the 117 amino acid sequence given in ABL06244; and (5) use of  
CC the polynucleotide for diagnosis and/or treatment of diseases caused by  
CC abnormal positioning of neural cells, and stimulation of association of  
CC reelin protein. The mouse reelin protein CR-50 epitope region has  
CC neuroprotective activity, and can be used in the diagnosis and treatment  
CC of cerebral disturbance due to an abnormal reelin gene and positioning  
CC of neurons. The present sequence represents a PCR primer for the mouse  
CC reelin protein CR-50 epitope region, which is used in an example from  
CC the present invention.

XX  
SQ Sequence 21 BP; 5 A; 5 C; 7 G; 4 T; 0 other;

Query Match 6.0%; Score 21; DB 24; Length 21;

Best Local Similarity 100.0%; Pred. No. 1.5;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGCAGTGTGGCACCACATG 21

Db 1 GAGCAGTGTGGCACCACATG 21

RESULT 10  
AAD22777  
ID AAD22777 standard; DNA; 20 BP.

XX  
AC AAD22777;

XX  
DT 26-FEB-2002 (first entry)

DE Mouse truncated reelin DNA existence detecting RL-11 RACE primer.

XX  
KW Mouse; reelin; F-spondin domain; CR-50 epitope; gene therapy; agyria;

XX RACE; rapid amplification of cDNA end; polymicrogyria; PCR primer;

XX ectopic gray matter; ss.

XX  
OS Mus musculus.

XX  
XX EP149844-A2.

XX  
XX 31-OCT-2001.

XX  
XX 11-APR-2001; 2001EP-0303411.

XX  
XX 11-APR-2000; 2000JP-0109954.

XX  
XX (RIKE ) RIKEN KK.

XX  
XX Mikoshiba K, Tabata H, Nakajima K;

XX  
XX WPI; 2002-019320/03.

PT Novel truncated Reelin protein containing F-spondin domain and CR-50  
PT recognition site of Reelin protein, but not having Reelin repeat site,  
PT useful to treat diseases including agyria due to abnormal neuron  
PT alignment -

XX  
XX Example 6; Page 8; 47pp; English.

CC The invention relates to a truncated Reelin protein comprising a  
CC F-spondin domain and a CR-50 recognition site but no reelin protein  
CC repeat site. Reelin is an essential molecule in developing a normal  
CC laminated structure of cerebrum. The truncated reelin protein and its  
CC DNA are useful for treating diseases including agyria, polymicrogyria,  
CC and ectopic gray matter due to abnormal neuronal alignment. Truncated  
CC reelin protein DNA is useful in gene therapy. The present sequence is  
CC a gene specific RACE (rapid amplification of cDNA end) PCR primer used  
CC for detecting mouse truncated reelin DNA existence.

XX  
SQ Sequence 20 BP; 4 A; 3 C; 7 G; 6 T; 0 other;

Query Match 5.7%; Score 20; DB 24; Length 20;

Best Local Similarity 100.0%; Pred. No. 4.8;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 191 CTGATTGATTCAGCTGGAG 210

Db 1 CTGATTGATTCAGCTGGAG 20

RESULT 11  
AAH19492  
ID AAH19492 standard; DNA; 4883 BP.

XX  
AC AAH19492;

XX  
DT 03-AUG-2001 (first entry)

XX  
DE Human coding sequence, SEQ ID 5.

XX  
KW Human; YS68 protein; early haematopoiesis; tumour; ss.

XX  
OS Homo sapiens.

CC F-spondin at the N terminus, a stretch of positively charged amino  
CC acids at the C terminus, and a series of eight internal repeats of  
CC 350-390 amino acids, each repeat containing two related sub-domains  
CC that flank a pattern of conserved cysteine residues known as an  
CC epidermal growth factor (EGF)-like motif. (I) has neuroprotective,  
CC neurotropic and antilipidemic activities, and can be used as a modulator  
CC of reelin-LDNR interaction. (I) is useful in screen for compounds that  
CC modulate reelin binding to an LDNR, in an assay system, where the assay  
CC system comprises a microplate array and an automated robotic  
CC microprocessor controlled system for adding and removing reagents to  
CC the microplate array. The compounds identified by the above screening  
CC method are useful as therapeutic agents to provide or alleviate a  
CC diverse spectrum of diseases including neurodegenerative disorders such  
CC as Alzheimer's disease, to facilitate neuronal regeneration after  
CC injury, to prevent or alleviate lipid metabolism diseases, to enhance  
CC cognitive functions and memory or to ameliorate other developmental  
CC disorders. The present sequence encodes human reelin, which is used in  
CC the exemplification of the present invention.

SO Sequence 11580 BP; 3014 A; 2696 C; 2753 G; 3116 T; 1 other;

Query Match Best Local Similarity 7.4%; Score 26; DB 24; Length 11580;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 208 GAGAAATTAGAGCCCTTCCATGT 233

DB 1067 GAGAAATTAGAGCCCTTCCATGT 1092

RESULT 7

AAS89484 ID AAS89484 standard; cDNA; 11632 BP.

XX AAS89484;

XX AC AAS89484;

XX DT 13-FEB-2002 (first entry)

XX DE DNA encoding novel human diagnostic protein #25288.

XX KM Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PR 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR P-PSDB; ABG25297.

XX PT New isolated polynucleotide and encoded polypeptides, useful in

XX PT diagnostics, forensics, gene mapping, identification of mutations

XX PT responsible for genetic disorders or other traits and to assess

XX PT biodiversity

XX Claim 1; SEQ ID No 25288; 103pp; English.

CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

SO Sequence 11632 BP; 3028 A; 2716 C; 2770 G; 3117 T; 1 other;

Query Match Best Local Similarity 7.4%; Score 26; DB 23; Length 11632;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 208 GAGAAATTAGAGCCCTTCCATGT 233

DB 1067 GAGAAATTAGAGCCCTTCCATGT 1092

RESULT 8

ABL40169/C ID ABL40169 standard; DNA; 22 BP.

XX ABL40169;

XX DT 21-MAY-2002 (first entry)

XX DE Mouse reelin protein CR-50 epitope region PCR primer SEQ ID NO:6.

XX KM Mouse; reelin protein CR-50 epitope region; elucidation; neuron;

XX KM cerebral disturbance; reelin protein; neuroprotective; PCR primer; ss.

XX OS Mus musculus.

XX PN JP2002017361-A.

XX PD 22-JAN-2002.

XX PR 04-JUL-2000; 2000JP-0202801.

XX PR 04-JUL-2000; 2000JP-0202801.

XX PA (RIKA) RIKAGAKU KENKYUSHO.

XX DR WPI; 2002-221707/28.

XX PT Reelin protein CR-50 epitope region, useful for diagnosis and treatment

XX PT of cerebral disturbance

XX PS Example 1; Page 6; 16pp; Japanese.

XX CC The present invention describes the mouse reelin protein CR-50 epitope

XX CC region, which contains the CR-50 antibody recognition site and is free

XX CC from F-spondin domains and repetitive sites. Also described are: (1) an

XX CC expression vector comprising a polynucleotide encoding a reelin protein

XX CC epitope region; (2) host cells with transfected the expression vector;

XX CC (3) polypeptides prepared by culture of the host cells; and (4)

XX CC polynucleotides comprising the 351 base sequence given in ABL40165 which

XX CC encodes the 117 amino acid sequence given in AB806244; and (5) use of

XX CC the polynucleotide for diagnosis and/or treatment of diseases caused by

XX CC abnormal positioning of neural cells, and stimulation of association of

XX CC reelin protein. The mouse reelin protein CR-50 epitope region has

XX CC neuroprotective activity, and can be used in the diagnosis and treatment

XX CC of cerebral disturbance due to an abnormal reelin gene and positioning

CC drugs or therapeutics for ischaemic diseases. AB199913 and AB199914  
 CC represent PCR primers for a mouse ischaemic condition related sequence,  
 CC which are used in the exemplification of the present invention.

XX Sequence 11673 BP; 2831 A; 2985 C; 2985 G; 2872 T; 0 other;

Query Match 100.0%; Score 351; DB 24; Length 11673;  
 Best Local Similarly 100.0%; Pred. No. 1.4e-166;  
 Matches 351; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GAGAGGTGTGGACCATCATGCGCATGCTGTCACCTCTCTGAGCGGTACGGCCT 60  
 DB 970 GAGCAGTGTGGACCATCATGCGCATGCTGTCACCTCTCTGAGCGGTACGGCCT 1029  
 OY 61 CGAGAGCTGACACACACATGCGCTGACACACACACATGCTGTCACATTTTCCATT 120  
 DB 1030 CGAGAGCTGACACACACATGCGCTGACACACACACATGCTGTCACATTTTCCATT 1089  
 OY 121 GGGTCAGGATCATGCTGATTTAGTTACTCTGACCCAGCATCTGTGTACGCCAAG 180  
 DB 1090 GGGTCAGGATCATGCTGATTTAGTTACTCTGACCCAGCATCTGTGTACGCCAAG 1149  
 OY 181 AACATACCGCTGATTTGGATTACGTCAGAGAAATTTAGAGCCCTTCCATGTAGCACA 240  
 DB 1150 AACATACCGCTGATTTGGATTACGTCAGAGAAATTTAGAGCCCTTCCATGTAGCACA 1209  
 OY 241 GTCATCCACATCTGTACCTCCCGAGAGAACCAAGGGAGAGCGTCATGTCACATGG 300  
 DB 1210 GTCATCCACATCTGTACCTCCCGAGAGAACCAAGGGAGAGCGTCATGTCACATGG 1269  
 OY 301 AAACAGCAGACCTGCGAGTGGGTGAGGTGTATGAGGCTGCTGGCCCTG 351  
 DB 1270 AAACAGCAGACCTGCGAGTGGGTGAGGTGTATGAGGCTGCTGGCCCTG 1320

## RESULT 5

ABN96939  
 ID ABN96939 standard; DNA; 11580 BP.

XX ABN96939;

DT 13-AUG-2002 (first entry)

DE Gene #3437 used to diagnose liver cancer.

KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;

KW metastatic liver tumour; cytostatic; expression profile; disease state;

KW disease progression; drug toxicity; drug efficacy; drug metabolism.

XX Homo sapiens.

PN WO200229103-A2.

PD 11-APR-2002.

XX 02-OCT-2001; 2001WO-US30589.

PR 02-OCT-2000; 2000US-237054P.

PA (GENE-) GENE LOGIC INC.

PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;

DR WPI; 2002-426119/45.

XX Diagnosing and detecting the progression of liver cancer,

PT hepatocellular carcinoma or metastatic liver tumor in a patient,

XX involves detecting the level of expression of two or more genes in a

XX liver tissue sample -  
 PS Claim 1; SEQ ID NO 3437; 298pp; English.  
 CC The invention relates to a novel method for diagnosing and detecting the

CC progression of liver cancer, hepatocellular carcinoma or metastatic liver  
 CC tumour in a patient, and differentiating metastatic liver cancer from  
 CC hepatocellular carcinoma in a patient, involving detecting the level of  
 CC expression of two or more genes represented in ABN93503-ABN97455 in a  
 CC tissue sample. The method of the invention has hepatotropic, and  
 CC cytostatic activity. The method is useful for diagnosing and detecting  
 CC the progression of liver cancer, hepatocellular carcinoma and metastatic  
 CC liver carcinoma in a patient. The method is useful for identifying  
 CC expression profiles which serve as useful diagnostic markers as well as  
 CC markers that can be used to monitor disease states, disease progression,  
 CC drug toxicity, drug efficacy and drug metabolism.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

SO Sequence 11580 BP; 3014 A; 2696 C; 2753 G; 3116 T; 1 other;

Query Match 7.4%; Score 26; DB 24; Length 11580;  
 Best Local Similarly 100.0%; Pred. No. 0.0041;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 208 GAGAAATTTAGAGCCCTTCCATGT 233  
 DB 1067 GAGAAATTTAGAGCCCTTCCATGT 1092

## RESULT 6

ABA92604  
 ID ABA92604 standard; CDNA; 11580 BP.

XX ABA92604;

DT 21-MAR-2002 (first entry)

DE Human reelin encoding CDNA SEQ ID NO:4.

KW Human; reelin; low density lipoprotein receptor; LDLR; neuroprotective;

KW extracellular glycoprotein; noctropic; anti-lipemic; Alzheimer's disease;

KW neurodegenerative disorder; neuronal regeneration; cognitive function;

KW lipid metabolism disease; memory; developmental disorder; gene; ss.

XX Homo sapiens.

PN US6323177-B1.

PD 27-NOV-2001.

XX 16-JUN-1999; 99US-0334220.

PR 16-JUN-1999; 99US-0334220.

PA (STUD-) ST JUDE CHILDREN'S RES HOSPITAL.

PI Curran T, D'Arcangelo G;

DR WPI; 2002-096596/13.

DR P-PSDB; ABB05007.

XX Novel composition useful for screening compounds that modulate Reelin

PT binding to low density lipoprotein receptor, comprising an isolated

PT Reelin polypeptide and low density lipoprotein receptor -

XX Example 1; Column 75-84; 45pp; English.

XX The present invention describes a composition (I) comprising an

CC isolated reelin protein (II) bound to an isolated low density lipoprotein

CC receptor (LDLR) (III). (II) is an extracellular glycoprotein of

CC approximately 385 kDa containing a small region of similarity with

RESULT 3  
 ABA92603  
 ID ABA92603 standard; cDNA: 11673 BP.  
 AC ABA92603;  
 XX  
 XX 21-MAR-2002 (first entry)  
 DE Mouse reelin encoding cDNA SEQ ID NO:3.  
 XX  
 XX Mouse; reelin; low density lipoprotein receptor; LDLR; neuroprotective;  
 KW extracellular glycoprotein; nootropic; antilipemic; Alzheimer's disease;  
 KW neurodegenerative disorder; neuronal regeneration; cognitive function;  
 KW lipid metabolism disease; memory; developmental disorder; gene; ss.  
 XX  
 OS Mus musculus.  
 XX  
 FH Key Location/Qualifiers.  
 FT 283..10668  
 CDS /\*tag= a  
 F /product= "mouse reelin"  
 XX  
 XX US6323177-B1.  
 XX  
 PD 27-NOV-2001.  
 XX  
 PF 16-JUN-1999; 99US-0334220.  
 XX  
 PR 16-JUN-1999; 99US-0334220.  
 XX  
 PA (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.  
 XX  
 PI Curran T, D'Arcangelo G;  
 DR WPI: 2002-096596/13.  
 DR P-PSDB: ABB05008.  
 XX  
 PT Novel composition useful for screening compounds that modulate Reelin  
 PT binding to low density lipoprotein receptor, comprising an isolated  
 PT Reelin polypeptide and low density lipoprotein receptor -  
 XX  
 XX Example 1: Column 65-74; 45pp; English.  
 XX  
 CC The present invention describes a composition (I) comprising an  
 CC isolated reelin protein (II) bound to an isolated low density lipoprotein  
 CC receptor (LDLR) (III). (II) is an extracellular glycoprotein of  
 CC approximately 385 kDa containing a small region of similarity with  
 CC corresponding at the N terminus, a stretch of positively charged amino  
 CC acids at the C terminus, and a series of eight internal repeats of  
 CC 350-390 amino acids, each repeat containing two related sub-domains  
 CC that flank a pattern of conserved cysteine residues known as an  
 CC epidermal growth factor (EGF)-like motif. (I) has neuroprotective,  
 CC nootropic and antilipemic activities, and can be used as a modulator  
 CC of reelin-LDLR interaction. (I) is useful in screen for compounds that  
 CC modulate reelin binding to an LDLR, in an assay system, where the assay  
 CC system comprises a microplate array and an automated robotic  
 CC microprocessor controlled system for adding and removing reagents to  
 CC the microplate array. The compounds identified by the above screening  
 CC method are useful as therapeutic agents to provide or alleviate a  
 CC diverse spectrum of diseases including neurodegenerative disorders such  
 CC as Alzheimer's disease, to facilitate neuronal regeneration after  
 CC injury, to prevent or alleviate lipid metabolism diseases, to enhance  
 CC cognitive functions and memory or to ameliorate other developmental  
 CC disorders. The present sequence encodes mouse (Mus musculus) reelin,  
 CC which is used in the exemplification of the present invention.  
 XX  
 XX Sequence 11673 BP; 2831 A; 2985 C; 2985 G; 2872 T; 0 other:

OY 1 GAGCAGTGTGGACACCATCATCTGGATGGCAATGCTGTACCTTCTGTGAGCCGTACGACCCT 60  
 Db 970 GAGCAGTGTGGACACCATCATCTGGATGGCAATGCTGTACCTTCTGTGAGCCGTACGACCCT 1029  
 OY 61 CGAGAGCTGTACCCACACATATGCTTGAACACACAGACATCTGTCTCCAGTTTTCATT 120  
 Db 1030 CGAGAGCTGTACCCACACATATGCTTGAACACACAGACATCTGTCTCCAGTTTTCATT 1089  
 OY 121 GGGTCAGATCATCTGTGATTTAGTTACTGTGACCCAGATACATCTGTGATACGCCAAG 180  
 Db 1090 GGGTCAGATCATCTGTGATTTAGTTACTGTGACCCAGATACATCTGTGATACGCCAAG 1149  
 OY 181 AACCATTACCGCTGATTTGATTCAGTTCAGTGGAGAAATTAGAGCCCTTCCAAATGTAGACACA 240  
 Db 1150 AACCATTACCGCTGATTTGATTCAGTTCAGTGGAGAAATTAGAGCCCTTCCAAATGTAGACACA 1209  
 OY 241 GTCATTCACATCTCTTACCTCCCGAGAAAGCCAAAGGGGAGAGGCGTCAGTTCCAGTGG 300  
 Db 1210 GTCATTCACATCTCTTACCTCCCGAGAAAGCCAAAGGGGAGAGGCGTCAGTTCCAGTGG 1269  
 OY 301 AAACAGGACAGCCTTGGAGTGGGTGAGTGTATGAGCCCTGCTGGGCCCTG 351  
 Db 1270 AAACAGGACAGCCTTGGAGTGGGTGAGTGTATGAGCCCTGCTGGGCCCTG 1320

RESULT 4  
 ABI99284  
 ID ABI99284 standard; cDNA: 11673 BP.  
 AC ABI99284;  
 XX  
 XX 07-MAR-2002 (first entry)  
 DE Mouse ischaemic condition related cDNA sequence SEQ ID NO:128.  
 XX  
 XX Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;  
 KW vasospastic ischemia; ischaemic condition; ischaemic disease; ss.  
 XX  
 OS Mus musculus.  
 PN WO200188188-A2  
 XX  
 PD 22-NOV-2001.  
 XX  
 PF 18-MAY-2001; 2001WO-JP04192.  
 XX  
 PR 18-MAY-2000; 2000JP-0145977.  
 XX  
 PA (UNYNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.  
 XX  
 PI Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;  
 DR WPI: 2002-034733/04.  
 DR P-PSDB; ABB57065.  
 XX  
 PT Examining the ischemic condition (e.g. occlusive ischemia) by measuring  
 PT expression levels of particular genes defined in the specification or  
 PT by determining the expression profile of a gene group comprising these  
 PT genes -  
 XX  
 PS Claim 2; Page 362-385; 2690pp; English.  
 XX  
 XX The present invention describes a method for examining ischaemic  
 CC conditions, comprising measuring the expression levels of particular  
 CC genes (I) in a test sample or determining the expression profile of a  
 CC gene group in the sample comprising genes selected from (I). The method  
 CC is useful for examining the ischaemic condition (e.g. compressive  
 CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring  
 CC expression levels of particular genes (ABI99202 to ABI99912, encoding  
 CC the protein sequences in ABB57020 to ABB57374) or by determining the  
 CC expression profile of a gene group comprising these genes. The  
 CC expression levels or expression profiles produced by these genes are  
 CC used as an indicator when screening for ischaemic condition-improving

DR WPI: 2002-221707/28.  
 DR P-PSDB; ABB06244.  
 XX Reelin protein CR-50 epitope region, useful for diagnosis and treatment  
 PT of cerebral disturbance -  
 PS Claim 5; Page 11; 16pp; Japanese.  
 XX The present sequence encodes the mouse reelin protein CR-50 epitope  
 CC region, which contains the CR-50 antibody recognition site and is free  
 CC from F-spondin domains and repetitive sites. Also described are: (1) an  
 CC expression vector comprising a polynucleotide encoding a reelin protein  
 CC epitope region; (2) host cells with transfected the expression vector;  
 CC (3) polypeptides prepared by culture of the host cells; and (4)  
 CC polynucleotides comprising the 351 base sequence given in ABL40165 which  
 CC encodes the 117 amino acid sequence given in ABB06244; and (5) use of  
 CC the polynucleotide for diagnosis and/or treatment of diseases caused by  
 CC abnormal positioning of neural cells, and stimulation of association of  
 CC reelin protein. The mouse reelin protein CR-50 epitope region has  
 CC neuroprotective activity, and can be used in the diagnosis and treatment  
 CC of cerebral disturbance due to an abnormal reelin gene and positioning  
 CC of neurons.  
 XX  
 SQ Sequence 351 BP; 86 A; 98 C; 92 G; 75 T; 0 other;  
 Query Match 100.0%; Score 351; DB 24; Length 351;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-166;  
 Matches 351; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GACGAGTGTGGACCATCATGATGCGATCTGTCACCTTCTGTGAGCCGTAGGCCCT 60  
 DB 1 GACGAGTGTGGACCATCATGATGCGATCTGTCACCTTCTGTGAGCCGTAGGCCCT 60  
 QY 61 CGAGAGCTGACCCACCATGCTGACCAACAACAGACATCTGCTCCAGTTTCCATT 120  
 DB 61 CGAGAGCTGACCCACCATGCTGACCAACAACAGACATCTGCTCCAGTTTCCATT 120  
 QY 121 GGGTCAGAGCATGTCGATTAGTACTGTACCCGACATCACTGTGTCAACGCCAAG 180  
 DB 121 GGGTCAGAGCATGTCGATTAGTACTGTACCCGACATCACTGTGTCAACGCCAAG 180  
 QY 181 AACAAATACCCCTGATGATGATTCAGCTGAGAGAAATTAAGACCCCTTCCAAATGAGCACA 240  
 DB 181 AACAAATACCCCTGATGATGATTCAGCTGAGAGAAATTAAGACCCCTTCCAAATGAGCACA 240  
 QY 241 GTCATCCACATCTCTGTAACCTCCCGAGAGCAAAAGGGAGAGCCGTGCAAGTCCAGTGG 300  
 DB 241 GTCATCCACATCTCTGTAACCTCCCGAGAGCAAAAGGGAGAGCCGTGCAAGTCCAGTGG 300  
 QY 301 AACAGAGACAGCCTGCGAGTGGGTGAGGTATGAGGCTGCTGGGCCCTG 351  
 DB 301 AACAGAGACAGCCTGCGAGTGGGTGAGGTATGAGGCTGCTGGGCCCTG 351  
 RESULT 2  
 AAD22754 standard; cDNA; 2745 BP.  
 XX AAD22754;  
 XX 26-FEB-2002 (first entry)  
 DE Mus musculus truncated reelin cDNA.  
 XX  
 XX Mouse: reelin; F-spondin domain; CR-50 epitope; gene therapy; agyria;  
 KM polymicrogyria; ectopic gray matter; ss.  
 OS Mus musculus.  
 XX  
 XX Key Location/Qualifiers  
 FH 283..2052  
 FT /tag= a  
 FT /product= "Mouse truncated reelin protein"

FT sig\_peptide 283..363  
 FT /tag= b  
 FT mat\_peptide 364..2049  
 FT /tag= c  
 FT /product= "Mature truncated reelin protein"  
 FT misc\_feature 284..849  
 FT /tag= d  
 FT /note= "Encodes F-spondin domain"  
 FT misc\_feature 970..1320  
 FT /tag= e  
 FT /note= "Encodes CR-50 epitope region"  
 XX  
 XX BP1149844-A2.  
 PD 31-OCT-2001.  
 XX  
 XX 11-APR-2001; 2001BP-0303411.  
 PR 11-APR-2000; 2000BP-0109954.  
 PR (Riken) RIKEN KK.  
 XX  
 PI Mikoshiba K, Tabata H, Nakajima K;  
 XX  
 DR WPI: 2002-019320/03.  
 DR P-PSDB; AAEI3606.  
 XX  
 PT Novel truncated Reelin protein containing F-spondin domain and CR-50  
 PT recognition site of Reelin protein, but not having Reelin repeat site,  
 PT useful to treat diseases including agyria due to abnormal neuron  
 PT alignment -  
 PS  
 PS Claim 10; Page 20-26; 47pp; English.  
 CC The invention relates to a truncated Reelin protein comprising a  
 CC F-spondin domain and a CR-50 recognition site but no reelin protein  
 CC repeat site. Reelin is an essential molecule in developing a normal  
 CC laminated structure of cerebrum. The truncated reelin protein and its  
 CC DNA are useful for treating diseases including agyria, polymicrogyria,  
 CC and ectopic gray matter due to abnormal neuronal alignment. Truncated  
 CC reelin protein DNA is useful in gene therapy. The present sequence is  
 CC a cDNA encoding Mus musculus truncated reelin protein.  
 XX  
 SQ Sequence 2745 BP; 661 A; 716 C; 714 G; 654 T; 0 other;  
 Query Match 100.0%; Score 351; DB 24; Length 2745;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-166;  
 Matches 351; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GACGAGTGTGGACCATCATGATGCGATCTGTCACCTTCTGTGAGCCGTAGGCCCT 60  
 DB 970 GACGAGTGTGGACCATCATGATGCGATCTGTCACCTTCTGTGAGCCGTAGGCCCT 1029  
 QY 61 CGAGAGCTGACCCACCATGCTGACCAACAACAGACATCTGCTCCAGTTTCCATT 120  
 DB 1030 CGAGAGCTGACCCACCATGCTGACCAACAACAGACATCTGCTCCAGTTTCCATT 1089  
 QY 121 GGGTCAGAGCATGTCGATTAGTACTGTACCCGACATCACTGTGTCAACGCCAAG 180  
 DB 1090 GGGTCAGAGCATGTCGATTAGTACTGTACCCGACATCACTGTGTCAACGCCAAG 1149  
 QY 181 AACAAATACCCCTGATGATGATTCAGCTGAGAGAAATTAAGACCCCTTCCAAATGAGCACA 240  
 DB 1150 AACAAATACCCCTGATGATGATTCAGCTGAGAGAAATTAAGACCCCTTCCAAATGAGCACA 1209  
 QY 241 GTCATCCACATCTCTGTAACCTCCCGAGAGCAAAAGGGAGAGCCGTGCAAGTCCAGTGG 300  
 DB 1210 GTCATCCACATCTCTGTAACCTCCCGAGAGCAAAAGGGAGAGCCGTGCAAGTGG 1269  
 QY 301 AACAGAGACAGCCTGCGAGTGGGTGAGGTATGAGGCTGCTGGGCCCTG 351  
 DB 1270 AACAGAGACAGCCTGCGAGTGGGTGAGGTATGAGGCTGCTGGGCCCTG 1320

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2002, 15:27:06 ; Search time 300 Seconds  
(without alignments)  
2634.838 Million cell updates/sec

Title: US-09-897-438b-1

Perfect score: 351  
Sequence: 1 gaggcgtgagcaccatcat.....atgagcctcgtggcgcctg 351

Scoring table: OLIGO\_NUC  
Gapco 60.0 , Gapext 60.0

Searched: 2185239 seqs, 112599159 residues

W size: 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

N\_Geneseq.101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	351	100.0	351	24	ABL40165
2	351	100.0	2745	24	AAD22754
3	351	100.0	11673	24	ABR92603
4	351	100.0	11673	24	ABR92604
5	26	7.4	11580	24	ABN96939
6	26	7.4	11580	24	ABN96939
7	26	7.4	11632	23	AAS89484
8	22	6.3	22	24	ABL40169
9	21	6.0	21	24	ABL40173

10	20	5.7	20	24	AAD22777	Mouse truncated re
11	19	5.4	4883	22	AAH19492	Human coding seque
12	19	5.4	7215	22	AAH19497	Human coding seque
13	19	5.4	7215	22	AAH19498	Human coding seque
14	18	5.1	18	24	ABL40174	Mouse reelin prote
15	18	5.1	975	21	AAV07071	Human interleukin
16	18	5.1	1035	21	AAV07071	Human interleukin
17	18	5.1	1074	18	AAV04440	SR345 coding seque
18	18	5.1	1260	20	AAV029202	Interleukin 6 rece
19	18	5.1	1466	19	AAV60296	Human IL-6 recepto
20	18	5.1	1545	21	AAV07073	Human interleukin-
21	18	5.1	1627	18	AAV97848	IL-6R/IL-6 fusion
22	18	5.1	2061	10	AAV90340	Human fusion poly
23	18	5.1	2066	14	AAV01346	Sequence encoding
24	18	5.1	2067	10	AAV0847	IL-6 receptor codi
25	18	5.1	2087	10	AAV0847	DNA contg. region
26	18	5.1	3319	17	AAV33369	Human secreted pro
27	18	5.1	3319	17	AAV31441	Interleukin-6 rece
28	18	5.1	3319	19	AAV60295	Human interleukin-
29	18	5.1	3319	21	AAV21364	Human low adenosin
30	18	5.1	3319	21	AAV35242	Human adenosine re
31	18	5.1	3477	21	AAV84527	Human CDNA differe
32	18	5.1	3507	21	AAV09046	Fusion polypeptide
33	18	5.1	4513	21	AAV21365	Fusion polypeptide
34	18	5.1	4873	21	AAV35243	Human low adenosin
35	18	5.1	9720	24	ABJ56176	Human adenosine re
36	18	5.1	198285	24	ABR84699	Hordeum vulgare va
37	18	5.1	198285	24	ABN97319	Human CDNA differe
38	17	4.8	22	21	AAC63485	Gene #3817 used to
39	17	4.8	47	18	AAT90825	beta-actin gene PC
40	17	4.8	325	22	AAV36445	Bacillus stearothe
41	17	4.8	335	18	AAV13260	Human musculuske
42	17	4.8	429	24	ABN95973	N. tabacum strain
43	17	4.8	559	22	ABV63804	Gene #2471 used to
44	17	4.8	559	22	ABV63804	Human foetal liver
45	17	4.8	559	22	AAK12314	Probe #9458 for ge
						Human brain expres

## ALIGNMENTS

RESULT 1	ABL40165	standard; DNA: 351 BP.
ID	ABL40165	
AC	ABL40165:	
XX	XX	
DT	21-MAY-2002	(first entry)
XX	XX	
DE	Mouse reelin protein CR-50 epitope region encoding DNA SEQ ID NO:1.	
XX	XX	
KW	Mouse; reelin protein CR-50 epitope region; elucidation; neuron;	
KM	cerebral disturbance; reelin protein; neuroprotective; gene; ds.	
XX	XX	
OS	Mus musculus.	
XX	XX	
FH	Key	Location/Qualifiers
FT	CDS	1..351
FT		/*tag= a
FT		/partial
FT		/product= "reelin protein CR-50 epitope region"
FT		/note= "no start or stop codons given"
PN	JP2002017361-A.	
XX	XX	
PD	22-JAN-2002.	
XX	XX	
PF	04-JUL-2000: 2000JP-0202801.	
XX	XX	
PR	04-JUL-2000: 2000JP-0202801.	
XX	XX	
PA	(RIKA ) RIKAGAKU KENKUSHO.	
XX	XX	



Query Match 4.8%; Score 17; DB 2; Length 47;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 267 GGAAGCCAAAGGGGAGA 283  
Db 17 GGAAGCCAAAGGGGAGA 33

RESULT 13  
US-09-480-921B-9/C

; Sequence 9, Application US/09480921B  
; Patent No. 6387637  
; GENERAL INFORMATION:  
; APPLICANT: Levin, Joshua Z.  
; APPLICANT: Budziszewski, Gregory J.  
; APPLICANT: Potter, Sharon L.  
; APPLICANT: Weglich, Lynette M.  
; TITLE OF INVENTION: Herbicide Target Genes and Methods  
; CURRENT FILING DATE: 2000-01-11  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 9  
; LENGTH: 1434  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1434)  
US-09-480-921B-9

Query Match 4.8%; Score 17; DB 4; Length 1434;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 195 TTGGATTGAGCTGAGA 211  
Db 38 TTGGATTGAGCTGAGA 22

RESULT 14  
US-09-480-921B-28/C  
; Sequence 28, Application US/09480921B  
; Patent No. 6387637  
; GENERAL INFORMATION:  
; APPLICANT: Levin, Joshua Z.  
; APPLICANT: Budziszewski, Gregory J.  
; APPLICANT: Potter, Sharon L.  
; APPLICANT: Weglich, Lynette M.  
; TITLE OF INVENTION: Herbicide Target Genes and Methods  
; FILE REFERENCE: PB/5-30780A  
; CURRENT APPLICATION NUMBER: US/09/480,921B  
; CURRENT FILING DATE: 2000-01-11  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 28  
; LENGTH: 1434  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(1434)  
; OTHER INFORMATION: encodes SEQ ID NO:29  
US-09-480-921B-28

Query Match 4.8%; Score 17; DB 4; Length 1434;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 195 TTGGATTGAGCTGAGA 211

Db 38 TTGGATTGAGCTGAGA 22

RESULT 15  
US-09-108-010B-11  
; Sequence 11, Application US/09108010B  
; Patent No. 6362399  
; GENERAL INFORMATION:  
; APPLICANT: ANTHONY JOHN KINNEY  
; GARY MICHAEL FADER  
; TITLE OF INVENTION: SUPPRESSION OF SPECIFIC  
; CLASSES OF SOYBEAN SEED  
; PROTEIN GENES  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY  
; STREET: 1007 MARKET STREET  
; CITY: WILMINGTON  
; STATE: DELAWARE  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 19898  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: DISKETTE, 3.50 INCH  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: MICROSOFT WINDOWS 95  
; SOFTWARE: MICROSOFT WORD 7.0A  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/108,010B  
; FILING DATE: 30-Jun-1998  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/019,940  
; FILING DATE: JUNE 14,1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: LYNNE M. CHRISTENBURY  
; REGISTRATION NUMBER: 30,971  
; REFERENCE/DOCKET NUMBER: BB-1071-A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 302-773-0164  
; TELEFAX: 302-773-0164  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1488 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; SEQUENCE DESCRIPTION: SEQ ID NO: 11:  
US-09-108-010B-11

Query Match 4.8%; Score 17; DB 4; Length 1488;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 261 CCCCGAAGCAAG 277  
Db 898 CCCCGAAGCAAG 914

Search completed: November 6, 2002, 18:13:18  
Job time : 78 secs

**THIS PAGE BLANK (USPTO)**

Sequence 2, Application US/08795473B  
Patent No. 6217858  
GENERAL INFORMATION:  
APPLICANT: Galun, Elhan  
APPLICANT: Nahot, Orit  
TITLE OF INVENTION: A Pharmaceutical Composition for Treating  
TITLE OF INVENTION: Hepatitis B Virus (HBV) Infection  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Davidson, Davidson and Kappel, LLC  
STREET: 1140 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: MS-DOS EDITOR  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/795,473B  
FILING DATE: 11-FEB-1997  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Davidson, Clifford M.  
REGISTRATION NUMBER: 32,728  
REFERENCE/DOCKET NUMBER: 963.1007  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)-997-1028  
TELEFAX: (212)-997-1037  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3319 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
US-08-795-473B-2

Query Match 5.1%; Score 18; DB 4; Length 3319;  
Best Local Similarity 100.0%; Pred. No. 3.7;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 CCAGCATCAGTGTGTCTCAT 172  
|||||  
Db 1256 CCAGCATCAGTGTGTCTCAT 1273

RESULT 11  
US-09-439-856-2  
Sequence 2, Application US/09439856  
Patent No. 6410009  
GENERAL INFORMATION:  
APPLICANT: Galun, Elhan  
APPLICANT: Nahot, Orit  
TITLE OF INVENTION: A Pharmaceutical Composition for Treating  
TITLE OF INVENTION: Hepatitis B Virus (HBV) Infection  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Davidson, Davidson and Kappel, LLC  
STREET: 1140 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: MS-DOS EDITOR  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/439,856  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/795,473  
FILING DATE: 11-FEB-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Davidson, Clifford M.  
REGISTRATION NUMBER: 32,728  
REFERENCE/DOCKET NUMBER: 963.1007  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)-997-1028  
TELEFAX: (212)-997-1037  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3319 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
US-09-439-856-2

Query Match 5.1%; Score 18; DB 4; Length 3319;  
Best Local Similarity 100.0%; Pred. No. 3.7;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 CCAGCATCAGTGTGTCTCAT 172  
|||||  
Db 1256 CCAGCATCAGTGTGTCTCAT 1273

RESULT 12  
US-08-642-684-13  
Sequence 13, Application US/08642684  
Patent No. 5834253  
GENERAL INFORMATION:  
APPLICANT: HONG, GUO FAN  
APPLICANT: FENG, ZHAI  
TITLE OF INVENTION: A NEW DNA POLYMERASE WITH PROOF-READING  
TITLE OF INVENTION: 3'-5' EXONUCLEASE ACTIVITY  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CUSHMAN, DARBY & CUSHMAN  
STREET: 1100 NEW YORK AVENUE, N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/642,684  
FILING DATE: 03-MAY-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: CHAPIN, MARIANA K.  
REGISTRATION NUMBER: 35,843  
REFERENCE/DOCKET NUMBER: 4694/219502  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-861-3711  
TELEFAX: 202-822-0944  
TELEX: 6714627 CUSH  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 47 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-642-684-13

NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Davidson, Davidson and Kappel, LLC  
STREET: 1140 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: MS-DOS EDITOR  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/795,473B  
FILING DATE: 11-FEB-1997  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Davidson, Clifford M.  
REGISTRATION NUMBER: 32,728  
REFERENCE/DOCKET NUMBER: 963.1007  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)-997-1028  
TELEFAX: (212)-997-1037  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1486 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
US-08-795-473B-3

Query Match  
Best Local Similarity 5.1%; Score 18; DB 4; Length 1486;  
Matches 18; Conservative 0; Pred. No. 3.6;  
Mismatches 0; Indels 0; Gaps 0;

OY 155 CCAGCATCAGTGTGTCAT 172  
Db 870 CCAGCATCAGTGTGTCAT 887

RESULT 7  
US-09-439-856-3  
Sequence 3, Application US/09439856  
Patent No. 641009  
GENERAL INFORMATION:  
APPLICANT: Galun, Elchan  
APPLICANT: Nahot, Orit  
TITLE OF INVENTION: A pharmaceutical Composition for treating  
TITLE OF INVENTION: Hepatitis B Virus (HBV) Infection  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Davidson, Davidson and Kappel, LLC  
STREET: 1140 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: MS-DOS EDITOR  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/439,856  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/795,473  
FILING DATE: 11-FEB-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Davidson, Clifford M.

REGISTRATION NUMBER: 32,728  
REFERENCE/DOCKET NUMBER: 963.1007  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)-997-1028  
TELEFAX: (212)-997-1037  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1486 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
US-09-439-856-3

Query Match  
Best Local Similarity 5.1%; Score 18; DB 4; Length 1486;  
Matches 18; Conservative 0; Pred. No. 3.6;  
Mismatches 0; Indels 0; Gaps 0;

OY 155 CCAGCATCAGTGTGTCAT 172  
Db 870 CCAGCATCAGTGTGTCAT 887

RESULT 8  
5171840-1  
Patent No. 5171840  
APPLICANT: KISHIMOTO, TADAMITSU  
TITLE OF INVENTION: RECEPTOR PROTEIN FOR HUMAN B CELL  
STIMULATORY FACTOR-2  
NUMBER OF SEQUENCES: 11  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/298,694  
FILING DATE: 19-JAN-1989  
SEQ ID NO: 1  
LENGTH: 2061

Query Match  
Best Local Similarity 5.1%; Score 18; DB 6; Length 2061;  
Matches 18; Conservative 0; Pred. No. 3.6;  
Mismatches 0; Indels 0; Gaps 0;

OY 155 CCAGCATCAGTGTGTCAT 172  
Db 1065 CCAGCATCAGTGTGTCAT 1082

RESULT 9  
5480796-1  
Patent No. 5480796  
APPLICANT: KISHIMOTO, TADAMITSU  
TITLE OF INVENTION: ANTIBODIES AGAINST THE RECEPTOR PROTEIN  
FOR HUMAN B CELL STIMULATORY FACTOR-2  
NUMBER OF SEQUENCES: 8  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/907,650  
FILING DATE: 02-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 298,694  
FILING DATE: 19-JAN-1989  
SEQ ID NO: 1  
LENGTH: 2061

Query Match  
Best Local Similarity 5.1%; Score 18; DB 6; Length 2061;  
Matches 18; Conservative 0; Pred. No. 3.6;  
Mismatches 0; Indels 0; Gaps 0;

OY 155 CCAGCATCAGTGTGTCAT 172  
Db 1065 CCAGCATCAGTGTGTCAT 1082

RESULT 10  
US-08-795-473B-2

## RESULT 2

US-09-334-220-4  
; Sequence 4, Application US/09334220  
; Patent No. 6323177  
; GENERAL INFORMATION:  
; APPLICANT: St. Jude's Children's Research Hospital  
; APPLICANT: Curran, Thomas  
; APPLICANT: D'Arcangelo, Gabriella  
; TITLE OF INVENTION: INTERACTION OF REELIN WITH VERY LOW  
; TITLE OF INVENTION: DENSITY LIPOPROTEIN (VLDL) RECEPTOR FOR SCREENING AND  
; FILE REFERENCE: 2427/0F704  
; CURRENT APPLICATION NUMBER: US/09/334,220  
; CURRENT FILING DATE: 1999-06-16  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FASTSEQ for Windows Version 3.0  
; SEQ ID NO: 4  
; LENGTH: 11580  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-334-220-4

Query Match 7.4%; Score 26; DB 4; Length 11580;  
Best Local Similarity 100.0%; Pred. No. 0.00024;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GAGAAATTGAGCCCTTCCATGT 233  
|||||

Db 1067 GAGAAATTGAGCCCTTCCATGT 1092  
|||||

## RESULT 3

US-08-627-151A-15  
; Sequence 15, Application US/08627151A  
; Patent No. 5866341  
; GENERAL INFORMATION:  
; APPLICANT: SPINELLA, Dominic  
; APPLICANT: BECHERER, Kathleen  
; APPLICANT: BROWN, Steven  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: SCREENING DRUG LIBRARIES  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Gen-Probe Incorporated  
; STREET: 10210 Genetic Center Drive  
; CITY: San Diego  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92121

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/627,151A  
; FILING DATE: 03-APR-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:

ATTORNEY/AGENT INFORMATION:  
; NAME: Fisher, Carlos A  
; REGISTRATION NUMBER: 36,510  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-410-8926  
; TELEFAX: 619-410-8928  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1074 base pairs

; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-627-151A-15

Query Match 5.1%; Score 18; DB 2; Length 1074;  
Best Local Similarity 100.0%; Pred. No. 3.6;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 CCAGCATCACTGTGTCT 172  
|||||

Db 819 CCAGCATCACTGTGTCT 836  
|||||

## RESULT 4

5171840-8  
; Patent No. 5171840  
; APPLICANT: KISHIMOTO, TADAMITSU  
; TITLE OF INVENTION: RECEPTOR PROTEIN FOR HUMAN B CELL  
; STIMULATORY FACTOR-2  
; NUMBER OF SEQUENCES: 11  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/298,694  
; FILING DATE: 19-JAN-1989  
; SEQ ID NO: 8  
; LENGTH: 1404  
5171840-8

Query Match 5.1%; Score 18; DB 6; Length 1404;  
Best Local Similarity 100.0%; Pred. No. 3.6;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 CCAGCATCACTGTGTCT 172  
|||||

Db 819 CCAGCATCACTGTGTCT 836  
|||||

## RESULT 5

5480796-8  
; Patent No. 5480796  
; APPLICANT: KISHIMOTO, TADAMITSU  
; TITLE OF INVENTION: ANTIBODIES AGAINST THE RECEPTOR PROTEIN  
; FOR HUMAN B CELL STIMULATORY FACTOR-2  
; NUMBER OF SEQUENCES: 8  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/907,650  
; FILING DATE: 02-JUL-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 298,694  
; FILING DATE: 19-JAN-1989  
; SEQ ID NO: 8  
; LENGTH: 1404  
5480796-8

Query Match 5.1%; Score 18; DB 6; Length 1404;  
Best Local Similarity 100.0%; Pred. No. 3.6;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 CCAGCATCACTGTGTCT 172  
|||||

Db 819 CCAGCATCACTGTGTCT 836  
|||||

## RESULT 6

US-08-795-473B-3  
; Sequence 3, Application US/08795473B  
; Patent No. 6217858  
; GENERAL INFORMATION:  
; APPLICANT: Galun, Elhan  
; APPLICANT: Nahot, Orli  
; APPLICANT: Blum, Herbert E.  
; TITLE OF INVENTION: A Pharmaceutical Composition for Treating  
; Hepatitis B Virus (HBV) Infection

